UNIVERSITY OF ILLINOIS

MAY 1 1961

THIS IS TO CERTIFY THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Steven Joseph Peter

ENTITLED: Barking Up the Wrong Family Tree? Greenberg's Method of Mass Comparison and the Genetic Classification of Languages

IS APPROVED BY ME AS FULFILLING THIS PART OF THE REQUIREMENTS FOR THE DEGREE OF Bachelor of Arts in Linguistics

Instructor in Charge

HEAD OF DEPARTMENT OF Linguistics
Barking Up the Wrong Family Tree?

Greenberg's Method of Mass Comparison and the Genetic Classification of Languages

By

Steven Joseph Peter

Thesis

for the
Degree of Bachelor of Arts
in
Linguistics

College of Liberal Arts and Sciences
University of Illinois
Urbana, Illinois

1991
# Table of Contents

0.0 INTRODUCTION ......................................................................................... 1

1.0 COMPARATIVE HISTORICAL LINGUISTICS ........................................ 2
  1.0.1 PRINCIPLES OF COMPARATIVE LINGUISTICS ............................ 3
  1.0.2 STRUCTURE AND CONVERGENCE .......................................... 5
  1.0.3 BASIC VOCABULARY ................................................................. 6
  1.0.4 CORRESPONDENCES BETWEEN LANGUAGES ............................ 7

1.1 BORROWING ......................................................................................... 8
  1.1.1 TIME DEPTH OF BORROWINGS .............................................. 9
  1.1.2 PROBABILITY OF BORROWING .............................................. 10
  1.1.3 METHODS OF DETECTING BORROWINGS ............................... 11

1.2 CHANCE SIMILARITIES ........................................................................ 11
  1.2.1 CHANCE AND THE SHORT MORPHEME .................................. 12
  1.2.2 CHANCE IN THE PROTOLANGUAGE ....................................... 13

1.3 ONOMATOPOEIA ................................................................................ 14

1.4 NURSERY LANGUAGE ......................................................................... 15

1.5 FILTERING OUT PROBLEMATIC ITEMS .......................................... 16
  1.5.1 THE "ORTHODOX" GREENBERG ............................................ 16

1.6 CRITERIA FOR DECISIONS ................................................................. 17
  1.6.1 VISUAL CORRESPONDENCES ................................................ 18
  1.6.2 RECURRENT CORRESPONDENCES ........................................ 18
  1.6.3 SHARED ABERRANCIES ............................................................ 19
  1.6.4 RECONSTRUCTION ................................................................. 20

2.0 AMERINDIANIST TRADITION .............................................................. 21
  2.0.1 THE ATTEMPTS AT CLASSIFICATION ...................................... 22
  2.0.2 LEXICOSTATISTICS ................................................................. 24
4.7.2 VOWEL CORRESPONDENCES BETWEEN ENGLISH AND HINDI 62
4.7.3 ETYMOLOGICAL CONSIDERATIONS 63

4.8 ENGLISH-HINDI-FINNISH ................................................................. 67

4.8.1 CONSONANT CORRESPONDENCES AMONG ENGLISH, HINDI AND FINNISH 68
4.8.2 VOWEL CORRESPONDENCES AMONG ENGLISH, HINDI AND FINNISH 68
4.8.3 ETYMOLOGICAL CONSIDERATIONS 69

4.9 DISCUSSION OF CORRESPONDENCES ............................................. 72

5.0 IMPLICATIONS FOR GREENBERG’S APPROACH ................................ 74

APPENDIX

84 CONCEPTS SELECTED FROM GREENBERG’S (1987) AMERIND DICTIONARY 76

SOURCES 81
ACKNOWLEDGMENTS

I would like to take this opportunity to thank IRJA ALHO for her aid in selecting the Finnish words, as well as searching for Finnish etymological materials for me. Her help is greatly appreciated.

To Professor HANS HENRICH HOCK, many thanks for his guidance, keen eye for detail, and patience, especially as my procrastination got the better of me at times.

Special thanks also are due for ALEX FRANCIS for listening to numerous drafts of this paper and for putting up with Professor Hock and me in our attempt to sift through the central problems in the paper.

GEOFFREY B. MUCKENHIRN and JOAN GOOD ERICKSON read parts of earlier drafts of this paper, and their comments were of great assistance to me. All errors of interpretation, of course, remain solely my own.
0.0 INTRODUCTION

Throughout the history of comparative linguistics, attempts have been made to come to terms with the difficulties in explaining the relationship between languages. Many problems have surfaced which must be dealt with in the exposition of the relationship of languages. For example, chance similarities prevail in natural language, onomatopoeia and nursery language are widespread, and borrowing makes its contribution to nearly every vocabulary. One of the earliest attempts at a solution was the comparative method, which devised means to deal with these problems. Recently, however, several linguists have proposed different methodologies to attempt to trace genealogical relationships back further than was previously done. For instance, the Nostraticists have begun to attempt to push back established time depth for distantly related languages, using a somewhat modified traditional approach. Making perhaps the strongest break with tradition, the Stanford linguist Joseph Greenberg has developed a methodology known as mass or multilateral comparison. Instead of comparing large numbers of vocabulary items across a few languages at a time, Greenberg compares a limited core of words across many languages, which he claims yields better results.

The method has not been without controversy, however. It has been claimed (Fodor 1982, Chafe 1987, Campbell, 1988, Matisoff 1990) that the method is unable to filter out chance similarities and the effects of borrowing. Greenberg has countered that because his
method considers many languages at a time, it will produce reliable results.

Clearly, the two camps have not come to agreement. An empirical study evaluating the cogency of the method is needed, which I provide in this paper. In my study, I apply Greenberg's method to an empirical test case, in which I compare English, Hindi, and Finnish. Because the basic relationship and histories of these languages are known, this study will serve as a check on the validity of Greenberg's method.

1.0 COMPARATIVE HISTORICAL LINGUISTICS

The Sanscrit language, whatever be its antiquity, is of a wonderful structure, more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly be produced by accident; so strong, indeed, that no philologer could examine all three, without believing that they have sprung from some common source, which, perhaps, no longer exists ... (Jones, 1786 in Hock 1986: 556)

The oft-quoted statement of Sir William Jones illustrates quite nicely the origins of comparative linguistics. Jones was led by structural and lexical similarities in three ancient languages to the belief that they were genetically related, i.e., that they had descended from a common source, and that there was (possibly) some sort of protolanguage that was mother to all three. Furthermore, the resemblances among the languages showed that they are genetically
related because the similarities could not "possibly be produced by accident".

These points, genetic relationship suggested by resemblances not produced by accident and the existence of a protolanguage to be reconstructed, have formed the backbone of comparative linguistics. During the last half of the 18th century and the first half of the 19th century, the main focus of comparative linguistics was centered around the Indo-European languages. In the process of establishing the genetic unity of Indo-European, as well as Finno-Ugric and other language families, comparative linguists came to develop a methodology known as the comparative method.

1.0.1 Principles of Comparative Linguistics

Meillet 1924 brings forward many of the notions with which comparative linguists work implicitly on a day to day basis. The basic concepts expounded upon in Meillet have been expanded and reworked, for example, by Greenberg 1957 and Hock 1986. The considerations brought forth by these linguists will weigh heavily on the later discussion in this paper. Thus, it will be useful to expound upon them in some detail here, following mainly Meillet, but also drawing on Greenberg and Hock, and supplemented by Matisoff.

Meillet begins by defining two types of comparison. The first is comparing in order to draw universal laws. The second, and for our purpose more relevant, type of comparison is made to find historical information. While, according to Meillet, both types are legitimate,
finding universals uncovers nothing more. It does not offer insight into the history of whatever is being investigated, which in this case is language.

Meillet also says that the arbitrary nature of the linguistic sign makes comparative linguistics possible. If there were a natural connection between words and their meaning, then the comparative method could extrapolate only general universals but not historical information.

Let us examine a list of words presented by Meillet (p. 15; diez is incorrectly cited as dies):

<table>
<thead>
<tr>
<th>French</th>
<th>Italian</th>
<th>Spanish</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>un, une</em></td>
<td><em>uno, una</em></td>
<td><em>uno, una</em></td>
<td>'one'</td>
</tr>
<tr>
<td><em>deux</em></td>
<td><em>due</em></td>
<td><em>dos</em></td>
<td>'two'</td>
</tr>
<tr>
<td><em>trois</em></td>
<td><em>tre</em></td>
<td><em>tres</em></td>
<td>'three'</td>
</tr>
<tr>
<td><em>quatre</em></td>
<td><em>quattro</em></td>
<td><em>cuatro</em></td>
<td>'four'</td>
</tr>
<tr>
<td><em>cinq</em></td>
<td><em>cinque</em></td>
<td><em>cinco</em></td>
<td>'five'</td>
</tr>
<tr>
<td><em>six</em></td>
<td><em>sei</em></td>
<td><em>seis</em></td>
<td>'six'</td>
</tr>
<tr>
<td><em>sept</em></td>
<td><em>sette</em></td>
<td><em>siete</em></td>
<td>'seven'</td>
</tr>
<tr>
<td><em>huit</em></td>
<td><em>otto</em></td>
<td><em>ocho</em></td>
<td>'eight'</td>
</tr>
<tr>
<td><em>neuf</em></td>
<td><em>nuove</em></td>
<td><em>nueve</em></td>
<td>'nine'</td>
</tr>
<tr>
<td><em>dix</em></td>
<td><em>dieci</em></td>
<td><em>diez</em></td>
<td>'ten'</td>
</tr>
<tr>
<td><em>vingt</em></td>
<td><em>venti</em></td>
<td><em>veinte</em></td>
<td>'twenty'</td>
</tr>
<tr>
<td><em>trente</em></td>
<td><em>trenta</em></td>
<td><em>treinta</em></td>
<td>'thirty'</td>
</tr>
<tr>
<td><em>quarante</em></td>
<td><em>quaranta</em></td>
<td><em>cuarenta</em></td>
<td>'forty'</td>
</tr>
<tr>
<td><em>cent</em></td>
<td><em>cento</em></td>
<td><em>ciento</em></td>
<td>'hundred'</td>
</tr>
</tbody>
</table>

One can clearly see a number of similarities between the languages. While the resemblance between *huit, otto,* and *ocho* may seem rather remote, other correspondences in these three languages establish the regularity of the correspondence. For example, note
the similarities between French *nuit*, Italian *notte*, and Spanish 
*noche* 'night'. Meillet (1924: 19) remarks that

These agreements, which cannot be explained by borrow-
ings from one language to another, presuppose a common 
origin. But it remains to interpret them in a systematic 
way: such is the object of comparative historical linguis-
tics.

What are the problems that comparative linguists have encountered 
in attempting to interpret such similarities?

1.0.2 STRUCTURE AND CONVERGENCE

Notice that the similarities are in words, that is, in the vocabu-
lar y. Attempts have also been undertaken to utilize structure to 
explain relationship between languages. There is a problem with explaining relationship using structure as the criterion, however. Structures in language are often of more limited scope, and thus likely to occur widely in languages, independent of genetic relationship.

Given subject, verb and object, there is a limited number of combinations possible (SVO, VSO, etc.), and these do not necessarily correlate with language relationships established on other evidence. For example, the modern Indo-European languages, a single genetic unit, when viewed in a simplified manner, exhibit three of the possible combinations; SOV in the East, SVO in most of Europe, VSO in insular Celtic (cf. Hock 1986: 561). As another example, English, Swedish, and German are all members of the Germanic family, yet
German has retained a more highly developed case system, whereas English and Swedish have all but eliminated it.

Furthermore, structure is readily diffused across genetic boundaries. This is known as convergence. For example, the basic sentence pattern in Finnish and Hungarian, located in Europe, is often more similar to their European neighbors than to the other Uralic languages. Likewise, the Indo-Iranian languages, part of the Indo-European family but located in close proximity to Dravidian and other non-Indo-European languages, have had strong structural influence from their neighbors.

Thus, general structural items within languages are not all that useful in determining genetic relationship. Instead, research has shown that prime consideration is to be given to vocabulary.

1.0.3 BASIC VOCABULARY

Having made the determination that vocabulary is more indicative of genetic relationship, comparative linguists began to sift through vocabularies to determine which items were most resistant to change. What developed out of this search was the concept of basic vocabulary.

For example, every society is expected to have some designation for 'to sleep' or 'to come', and these were present since time immemorial. On the other hand, not every society possessed writing or other technological and/or cultural innovations. Words referring to these items would not, then, be part of the basic vocabulary, and
would be more apt to be borrowed. Thus, the probability of borrowing of non-basic vocabulary is far greater.

1.0.4 CORRESPONDENCES BETWEEN LANGUAGES

Let us return to our list of numerals cited by Meillet. Given such a set of correspondences, Meillet says that one is led to posit that the numerals of French, Italian, and Spanish all stem from a common tradition. The tradition could be one of two types: either one or more borrowed from the others, or all three had a common origin. He points to the unlikelihood of *huit* being derived from *otto* or *ochô* via borrowing (and the other permutations of borrowing between the three languages) in order to discount the former possibility in favor of common origin.

The correspondences in the above example are quite transparent. When the languages in question have been separated by a greater length of time, the correspondences can be much more difficult to see. For example (Meillet 1924: 17-18), if one compares Sanskrit *d(u)vā*, Greek *duo*, Latin *duo*, and Armenian *erku*, the Armenian form does not seem to correspond with the others. However, if one looks at other examples, one can see that there is indeed a correspondence between Armenian *erk-* and *dwy-* of the other languages. “[T]hus, just as Greek has a root *dwi-* for the idea of ‘to fear’, Armenian has *erki-* (*erkiwl ‘fear’), and just as Greek has an old adjective *dwarón* for ‘long’, Armenian has *erkar* ... The agreement can thus be reduced to a general rule of correspondence: an old *dw-*
becomes Arm. *erk-.*\(^1\) Thus, a recurrent correspondence, even when it defies easy explanation, may be a strong indication of genetic relationship.

Not everything that corresponds between two languages, however, is the result of common heritage. Borrowing may affect the vocabulary in a language. Additional considerations which must be considered include chance similarities, onomatopoeia, and nursery language. The fundamental problem is how to detect and weed these out.

1.1 BORROWING

Detecting borrowing is a basic challenge to comparative linguists. Situations may arise where vocabulary in one language seems to be related to two other languages simultaneously. Consider the well-known example (as presented in Hock 1986: 559):

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>calf</td>
<td><em>Kalb</em></td>
<td>veau</td>
</tr>
<tr>
<td>veal</td>
<td><em>Kuh</em></td>
<td>bœuf</td>
</tr>
<tr>
<td>cow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beef</td>
<td>Schwein</td>
<td></td>
</tr>
<tr>
<td>swine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pork</td>
<td></td>
<td>porc</td>
</tr>
</tbody>
</table>

\(^1\) For a more detailed account of this correspondence, see Hock 1986. Here, Hock discusses the problematics in explaining such a correspondence. It is relatively straightforward, he shows, to arrive at an expected outcome of Armenian *tk-. It is, however, more problematic to explain how to get to the actual occurrence of *rk- from this form.
In contrast, comparison of other vocabulary in other semantic fields (twenty/zwanzig/vingt; eat/essen/manger; too/zul/trop) leads to the conclusion that English is closer to German and that the similarities with French in the above example result from borrowing.

1.1.1 Time Depth of Borrowings

More recent borrowings are more likely to be detected more easily than borrowings that entered the language at a time further in the past; cf. zeitgeist vs. bishop, both borrowings into English, the first from German (Zeitgeist) in the 19th century, the second from Latin (episcopus), borrowed in the Old English period. The first is still close to its German source, while the second, as a result of centuries of change, has become quite different from its Latin original.

Borrowings can occur at any time in the history of languages, including at the level of the protolanguage. When the latter occurs, the borrowed words are likely to appear in all of the daughter languages, and it becomes more difficult to sort out these items as borrowings, as they are so widespread across the languages in question. Notice, for example, that the West Germanic protolanguage borrowed quite early the Latin word tegula 'tile'. In the subsequent development of the languages, the words differentiated in the same way as native vocabulary, undergoing all the relevant sound shifts, etc. The modern words English tile, Dutch tegel, and German Ziegel, thus appear as native as English tide, Dutch
tijd ‘time’, and German Zeit ‘time’ (the latter three derived from Indo-European *dā[i]-). Borrowings occurring at a greater depth in time thus are much more difficult to detect.

1.1.2 PROBABILITY OF BORROWING

Meillet warns that the comparative linguist must remain critical at every juncture, so as not to be misled into believing that a word is inherited when, in fact, it is not:

The risk that a word is borrowed is always great, and the etymologist of an ancient or modern language who reasons as if the words to be explained had a priori every chance of being native exposes himself to frequent errors. (p. 51)

As examples of this, he observes that verbs and adjectives are more often native than substantives (e.g., the French verbs vivre ‘to live’, mourir ‘to die’, venir ‘to come’, and dormir ‘to sleep’, and the adjectives vif ‘vibrant’ and vieux ‘old’ reflect direct-descent Romance roots). Still, this is only a tendency. For instance, German has borrowed the verb schreiben ‘write’ from the Latin scribere, going so far in nativizing it as to make it a ‘strong’ verb (i.e., schreiben, schrieb, geschrieben). Conversely, French (as well as German) has many nouns which reflect inherited Indo-European roots. Thus, it is difficult to differentiate a priori native words and borrowings, especially not simply by stating that nouns must be suspect while verbs and adjectives can be approached more laxly.
1.1.3 METHODS OF DETECTING BORROWINGS

Greenberg 1957 further defines how to detect borrowing in his conclusion regarding the relationship between Thai, Sino-Tibetan, and Kadai/Malayo-Polynesian. He concludes that Thai must have borrowed from Sino-Tibetan (1957: 40):

The specific resemblances found with Sino-Tibetan languages always occur in forms found in Chinese, usually to the exclusion of other Sino-Tibetan languages. The specific form, even when found elsewhere, is always very close to Chinese. Moreover, the resemblances cluster in a few semantic spheres, the numerals from 2 to 10 and a few names of metals and domestic animals. In contrast, the Thai resemblances to the Kadai languages and Malayo-Polynesian tend to recur throughout the family, not just in some single language; are basic; do not concentrate in any particular semantic area; and exhibit an independence of form which excludes any particular Kadai or Malayo-Polynesian language as a source.

As Greenberg shows, borrowing tends to be concentrated in certain semantic fields (often more technological in nature). Thus, the evidence used in establishing genetic relationship should consist of a large number of correspondences distributed THROUGHOUT the vocabulary, especially BASIC vocabulary cf. the French-English correspondences in section 1.1 above.

1.2 CHANCE SIMILARITIES

Among the myriad combinations of sounds, there occur random similarities which may at first seem to prove a relationship, but in the end are merely humorous asides. As Matisoff (1990) observes,
nearly everyone has his or her own list of favorite chance similarities. For example, there is a word in both English and Persian, /baed/ (Meillet 1924), which has roughly the same form and meaning in both languages, yet the similarities can be attributed only to chance. (Note additionally that Didinga, a language spoken in Sudan, has *badh, meaning ‘bad’.) How does one avoid these chance similarities?

1.2.1 CHANCE AND THE SHORT MORPHEME

In determining genetic relationship, a necessary requirement is that the elements to be compared are to some extent substantial, as Meillet points out (1924: 52).

Thus linguists who work with small root elements and who often analyze roots to compare no more than fragments of roots thereby ruin in advance the proof which they seek ... A comparison which rests solely on one or even two root consonants is without value if it is not supported by very specific facts.

He writes that citing Sanskrit *vidhāvā, Old Church Slavonic *vidova, Old Prussian *widdewū, Gothic *widuwo, Irish *fedb, and Latin *uidua offers a great deal of proof, especially when recurrent sound correspondences are also given.

Meillet does not offer specific examples of short morpheme correspondences, but one might look at the morpheme –t in (Older) English. In Early Modern English, the suffix indicating second person singular of the verb was –st (e.g. thou say-est). It would be logical to conclude that English preserves a Germanic *-st form, yet we
know from the history of the language that this is not the case. The inherited form was -s. We are then left to explain the -t.

The other Indo-European languages also show a -t morpheme in the second person singular (cf. Sanskrit vētṛha ‘you know’, Greek oisthā ‘you know’, Old English wāst ‘you know’, and German wēβt ‘you know’). Is the -t of English sayest, then, the same as the -t of these forms?

It is not; the similarity is pure chance. The -t of sayest came about from quasi-affixation of the pronoun ḫu ‘thou’ in interrogation, which later spread to non-interrogative situations (e.g. *sages- ḫu, *sāgstu, sayest). The -t of the other forms reflects a Proto-Indo-European perfect ending *-t₁le.

According to Meillet, there is a grave danger in comparing not only short morphemes, but also ever dwindling portions of words to arrive at an acceptable etymology. "Whatever language is concerned, an etymology can be considered as proved only if a set of precise agreements establishes that the similarities of the words compared cannot be due to chance." (p. 54)

1.2.2 CHANCE IN THE PROTOLANGUAGE

There is also the possibility of chance similarities between protolanguages. Consider, for instance, that protolanguage A has a form *mak- ‘one’, while protolanguage B has, entirely through chance, *mag? ‘(index) finger’ (this is for illustrative purposes only, the protolanguages in question do not exist). The difficulty arises
that the daughter languages of both protolanguages will most likely show correspondences in this vocabulary item, pointing to a single source, when in reality two roots must be posited.

Therefore, the comparative linguist must insist on a large number of correspondences throughout the vocabulary, and the correspondences must be in substantial morphemes of more than merely a few phonemes. Only in doing so can the possibility that chance has played a role in the similarities between languages be held to a minimum.

1.3 ONOMATOPOEIA

As Meillet mentioned, the arbitrariness of the linguistic sign is a necessary prerequisite to comparative linguistics. At times, however, the connection between sound and meaning is more direct. The sound a gun makes is the same everywhere in the world. The designations in speech for this event are strikingly similar, although there is some variation (cf. English bang, German peng or puff, Spanish pen or pung, Russian [puf]).

In addition to this more direct connection, there is a cross-language tendency to associate certain phonetic elements with similar extra-linguistic events (cf. [l, i] in English drip, German triefen, for a higher-pitched occurrence of water falling into more water at a slow rate; or [ɔ] in English drop, German tropfen, referring to a lower-pitched drip). This can also be transferred to words dealing with
the other senses by synesthesia (cf. the high front vowel [i] related to a "sharp" visual image in *twinkle*).

The comparative linguist must be aware of items such as these, and where possible not include them as evidence of genetic relationship, unless, of course, they reflect other general correspondences. As in dealing with chance, demand must be placed on a large number of correspondences to eliminate the possibility of onomatopoeia as a source for similarities.

1.4 NURSERY LANGUAGE

A problem similar to onomatopoeia is nursery language. Babies all over the world babble. As adults hear these sounds, they often attempt to assign meaning to the sequences. This is not limited solely to English *dada, daddy* or *mama, mommy*; cf. Italian *papa, mama*, Hindi *bāp, mā*.

In addition to being limited to the semantic sphere of local kinship terms and a few cribside items\(^2\), nursery words tend to reduplicate (*da-da, na-na, ma-ma*), and tend to predominate in nasals, stops, and the vowel /a/, as these are the most recurrent sounds involved in babbling.

Examples of nursery language are usually easily detected, and must be eliminated from the evidence for genetic relationship. Along with the application of the cautionary rules formulated in the

---

\(^2\) Note however, that it may occur elsewhere; my early word for 'refrigerator' [ulululu] was in family usage for a number of years. Such words are unlikely to be passed into general usage, though.
sections regarding chance similarity, borrowing and onomatopoeia, these words should not present much of a problem.

1.5 FILTERING OUT PROBLEMATIC ITEMS

The strength of the comparative method is that it has devised methods to deal with the problematic items mentioned above. Meillet 1924 offers a rather nice summary of these approaches.

Perhaps the greatest caveat that Meillet presents in his work that will have a major effect on the analysis of Greenberg's approach is the following:

Whether it is a question of morphology, phonology, or vocabulary, the principle which must never be lost sight of is that comparisons are valid only to the degree that they are subjected to strict rules. The more freedom the linguist gives himself, the more arbitrary are his comparisons and the more precarious his proofs. (p. 57)

Thus, the basic premises of the comparative method are strict rules of correspondences, applied to many thousands of items in the languages being investigated.

1.5.1 THE "ORTHODOX" GREENBERG

Interestingly enough, in 1957, Greenberg professed many of the same points found in Meillet. He argued that there must be well-established sound/meaning correspondences, that longer forms are to be preferred over shorter ones, and he warns against borrowings and chance similarities, wide jumps in semantic resemblances, etc.

Greenberg writes:
Another factor bearing on the value of particular resemblances is semantic plausibility ... Semantic plausibility ... attaches to comparisons involving single-step, widely attested shifts in meaning, e.g. "moon" and "month". The more intermediate steps allowed, the larger the chance of obtaining form-meaning similarity, some of which may indeed stem from historical consideration. BUT THE GREATER THE METHODOLOGICAL LATITUDE PERMITTED, THE LESS PLAUSIBLE IS EACH INDIVIDUAL COMPARISON. (38, emphasis added)

See 3.1.1 below for a discussion of the controversy generated by Greenberg's methodological latitude vis-à-vis semantic plausibility in his later work.

1.6 CRITERIA FOR DECISIONS

It has been observed many times that two languages cannot be proven to be not related. Greenberg 1987 says that this is indeed true, but one may show relative degrees of relatedness. That is, one must show, for example, that Cherokee is more closely related to Mohawk than to Swahili in order to place Cherokee and Mohawk in a valid genetic unity (commonly called Iroquoian). See section 3.0 below for a further discussion of the implications of Greenberg's position. Note that a linguist may establish the degree of relatedness via a number of criteria of varying degrees of certainty.

Ultimately, comparative linguistics is based on making judgment calls. These decisions must be as well reasoned as possible, but there nearly always remains some disagreement. There is in essence a cline of certainty involved, beginning with visual corre-
spondences, followed by shared recurrent correspondences, aberrancies, and finally, reconstruction.

1.6.1 Visual Correspondences

There are cases in which one can see the relationship between languages, plain and simple (cf. Meillet’s list of numerals in French, Italian, and Spanish in section 1.0). This is likely to work only with languages related very closely. It is not likely to work for establishing genetic relationship among less-closely related languages, however.

1.6.2 Recurrent Correspondences

An important concept is that of systematic recurrent correspondences. Grimm (building upon work done by Rask) had postulated two Lautverschiebungsgesetze, or sound shift laws. The first law explained the shift from Indo-European to Germanic:

/p,t,k/ > /f,θ,x/

For example, Latin piscis : English fish; Latin pater : English father; Latin cornū : English horn.

/b,d,g/ > /p,t,k/

For example, Latin turbā ‘crowd’ : English thorp. Latin duō : English two, Latin genu : Old English ēnēo ‘knee’.

/bh,dh,gh/ > /b,d,g/

For example, Sanskrit bhṛater : English brother; Sanskrit dhēr : English do, IE *ghostis : English guest.
As the second law, he postulated a shift from Germanic to High German:

\[ /p,t,k/ > /pf,ts,kx/ \text{ or } /f,s,x/. \]

For example, English *pepper* : High German *Pfeffer*; English *tide* : High German *Zeit* /tsait/; English *make* : High German *machen* (In initial position, only in Swiss German, cf. /k xu:xən/).

These shifts represent diachronically what synchronically is a set of systematic recurrent correspondences. Grimm's Law pointed to a clear connection between Modern High German and Indo-European, and foreshadowed the possibility of reconstructing Proto-Indo-European.

However, one does not have to explain fully the diachronic details to establish relatedness. For example, it was known from the time of its initial proposal that there are several exceptions to Grimm's Law, which would have to be explained before reconstruction could occur. Nonetheless, the relationship between English and German, on the one hand, and Latin, on the other, was established through the application of the comparative method by finding a set of such systematic, recurrent correspondences, not through the reconstruction of Proto-Indo-European.

1.6.3 SHARED ABERRANCIES

Etymologies are not the only means that the comparative method has with which to establish the relationship between languages. As Meillet (1924: 41) notes, the notion of shared aberrancies is also
important. A transparent example is presented in Hock 1986: Both English and German have suppletion in the comparative and superlative forms for ‘good’. It is true that French also has suppletion for this comparison, but English and German share systematic, recurrent correspondences which French does not have. Significantly, such suppletion is not the normal manner of forming comparative and superlative forms (cf. English new, newer, newest; German neu, neuer, am neusten). Such an aberrancy is generally considered likely to indicate genetic relationship, not borrowing or chance:

   English: good  German: gut  French: bon
   better       besser       meilleur
   best         best-        le meilleur (p. 563).

1.6.4 RECONSTRUCTION

For many linguists, the most conclusive proof of the genetic relationship of languages is the reconstruction of a protolanguage. Compare the discussion in Hock 1986. As Hock notes, reconstruction in many cases is not possible because of the limited nature of available evidence. Hock suggests that this may be the case even for languages which we know are related. For instance, if we compare Sanskrit and Old English, the two languages are similar enough to establish relationship quite readily; reconstruction can be expected to be fairly easy. The modern descendents of these languages, however, Hindi and Modern English, have as a consequence of continued linguistic change become much more different, so much
so, in fact, that it may no longer be possible to reconstruct their common ancestor based on the evidence of these languages.

Since Greenberg's approach deals with languages that are quite remotely related — if they are related at all — the evidence is so limited that reconstruction is out of the question. The issue of reconstruction, therefore, need not further concern us here.

2.0 AMERINDIANIST TRADITION

The Amerindianist tradition presents a mixed picture vis-à-vis the Indo-European comparativist tradition. Most of the early attempts at the classification of the languages in the Americas revolved around typological, not genetic, considerations. The distinction that Meillet draws so poignantly between the two types of comparison are not to be found in early Amerindianist work. Durangeau 1819 noted that the languages in the Americas were polysynthetic. For a time, nearly every essay regarding these languages pointed out (with almost stock phrases) the similarity in typology from Greenland to Cape Horn.

Comparative linguists in the Americas faced certain problems that had not been encountered in Europe. The greatest initial problem was lack of vocabularies.

Through the use of the concept of basic vocabulary, which came from the Indo-Europeanist tradition, the comparative linguists in the Americas began to devise word lists. These lists were designed
to maximize the data collected during the short time available for fieldwork by providing specific information to be collected.

2.0.1 THE ATTEMPTS AT CLASSIFICATION

An early attempt at the genetic classification of the languages of North America (north of Mexico) was made by Powell 1891. On the basis of the newly collected word lists, he postulated 55 families, with 22 in California alone.

The next major revision to the schema of classification was made by Sapir, who attempted to reduce the number of families proposed by Powell. Sapir conducted most of his initial work via the accepted comparative method, and even published comparative studies of Uto-Aztecan and Athabaskan. Faced with the difficulties of larger-scale comparison and attempting to reduce the number of linguistic stocks, Sapir adopted a different approach. He based much of his later classificatory work on certain types of morphological and syntactic similarities. Thus, in discussing his proposed Hokan-Siouan stock (of six superstocks which he proposed to replace Powell’s 55), Sapir writes:

The Hokan-Siouan languages are prevailingy agglutinative, tend to use prefixes rather than suffixes for the more formal elements, particularly the pronominal elements of the verbs; distinguish active and static verbs; and make free use of compounding of stems and of nominal incorporation. (1929, reproduced in Haas 1978: 198)
Note, however, that this appraisal is more of a typological classification than a genetic one. This illustrates one of the problems in early classificatory work on Amerindian languages, the concerns were typological and not genetic, but the two were never carefully differentiated.

This brings us back to Meillet (1924), who makes the distinction clear in his very first paragraphs (cf. section 1.0 above). Note also the discussion above regarding the general unacceptability of structure alone for determining genetic relationship (section 1.0.2). Thus, Sapir’s structural criteria cannot be considered reliable for establishing linguistic relationship.

Most Amerindianist scholars found their way into linguistics via anthropology, but one influential scholar in this field, Bloomfield, entered through Germanic and Indo-European philology. While noting the great contributions made by Boas, who never personally did classificatory work, Bloomfield writes:

The native languages of our country have been studied by some very gifted men, but none had succeeded in putting this study on a scientific basis. The scientific equipment of linguists, on the other hand, contained few keen tools except the comparative method and this could not yet be here applied. (Bloomfield 1943)

Bloomfield attempted to bring the comparative method to bear on American languages. He wanted to show that even languages which did not have a large corpus of written material could be dealt with successfully by the comparative method. To accomplish this, he first undertook a great deal of fieldwork and descriptive work on several
Amerindian languages in order to have a solid foundation upon which to work. His article "On the sound system of Central Algonquian" (Bloomfield 1925) proved that the application of the comparative method to languages lacking in written records was indeed possible. Unfortunately, however, Bloomfield never became interested in undertaking larger groupings of Native American languages.

Two additional concepts stemming from the Amerindianist tradition are the notions of lexicostatistics and glottochronology. Both methods grew out of the use of word lists for dealing with languages in the Americas (cf. Powell 1891). While the two are similar, there are some important differences.

2.0.2 LEXICOSTATISTICS

Lexicostatistics was used originally as a method for subgrouping languages by making claims about the synchronic "distance" between languages. The critical features here are the concentration on synchronicity and the use of lexicostatistics as a first step only.

As pointed out above (§ 1.0.3), the word lists used for lexicostatistics developed out of the concept of basic vocabulary, on the assumption that basic vocabulary items are more resistant to change. The innovative assumption in lexicostatistics was that these items would be replaced at a constant rate in a given language or languages. Thus, languages which have a higher rate of correspondences are assumed to be more closely related than
languages which do not show as high a percentage of correspondences.

However, problems arose with the method. The original word list, composed by Swadesh, contained 200 basic-vocabulary concepts to be glossed into the languages under investigation. Further work showed that more and more words turned out to be problematic in one sphere or another and were dropped from the list. Instead, a 100-word list became widely used.

Furthermore, the assumed constant rate of change is not well supported empirically. Dyen 1975 shows that tabu can play a role in the rate of vocabulary replacement. As his example, he discusses a custom in Tahiti in which it is forbidden by social mores to use words that either formed a part of the name of deceased royalty or that sounded similar. This tabu, therefore, leads to large-scale replacement of vocabulary, including basic vocabulary, at an accelerated rate in comparison with other languages and/or societies. As the generational spacing of monarchs cannot be assumed to be a constant, this phenomenon will skew the results of a lexicostatistic study. It must be said, however, that Dyen is ultimately in favor of lexicostatistics and glottochronology, in spite of his own caveats.

2.0.3 GLOTTOCHRONOLOGY

Glottochronology is an attempt to apply the claims made by lexicostatistics to an investigation of time depth separating languages. Glottochronology is based on an analogy to the carbon dating
method in the physical sciences. Yet this view is flawed, as Hock (1976: 209) points out:

Radiocarbon, once it has undergone decay, ceases to be radiocarbon and can no longer undergo the same change. A determination of the ratio of radioactive carbon to normal carbon in a particular piece of (dead) wood thus will produce an accurate basis for the determination of the amount of decay in it (and the corresponding approximate time depth). The replacement of one lexical item, however, does not result in an item which can no longer undergo the same change. Rather, lexical replacement is repeatable.

We thus cannot be certain whether a given set of differences is the result of a single change or a series of changes (in one or both languages). Therefore, even if the rate of vocabulary replacement were constant, it is impossible to determine the number of changes and the corresponding time depth through mere data inspection.

However, Bergsland and Vogt (1962) made an empirical study of glottochronology. In comparing Modern Icelandic and three Norwegian dialects with Old Norse, Modern Georgian with Old Georgian and Mingrelian, and Modern Armenian with Old Armenian, they found that the rate of replacement was indeed not constant. For example, Modern Icelandic and Riksmål show a difference in retention rates of about 15%.

In this case, they state, the discrepancy can be easily explained: there are numerous borrowings that have effected the results. In this example, we know what the borrowings are; it is not, however,
so easy to pinpoint the borrowings in a 200-item word list of languages whose earlier history we do not know.

Bergsland, Vogt, and Dyen, then, have in effect destroyed the basic assumption that the rate of vocabulary chance is constant. If this is the case, glottochronology is unable to substantiate the claims that it had set out to make regarding the depth in time of the separation of languages from their respective families.

2.0.4 GREENBERG AND GLOTTOCHRONOLOGY

Greenberg apparently approves of glottochronology, and has indeed included an appendix on it in *Language in the Americas* (Greenberg 1987; hereafter also *LIA*). His primary purpose in doing so is to adapt glottochronology, developed for comparing only a few languages at a time, to multilateral comparison.

He sets out mathematically how much vocabulary will be preserved after a given length of time of separation (e.g., in looking at 20 languages, 2.2% of the vocabulary is recoverable after 20,000 years). Actually, he presumes a higher percentage, based on the expectation that vocabulary loss will be slower through time, as most of the unstable vocabulary is lost.

There remain serious doubts regarding the validity of lexicostatistics, which is the more conservative of the two approaches discussed here. Greenberg’s attempt to introduce glottochronology in conjunction with mass comparison casts further doubt on his overall approach.
3.0 GREENBERG’S APPROACH

Greenberg’s first attempt at classification using his still-developing approach of mass comparison was on the languages of Africa (Greenberg 1955). He began the series of articles, which were later collected into book form, by a brief introduction to his method. Subsequent criticism of his method led him to attach an appendix, in which he explains in somewhat more detail what he understands his new method to mean.

He begins by defending the idea that genetic relationships can be seen, plain and simple. He points to Sir William Jones (cf. 1.0) as an example, writing that Jones saw the relationship of three of the Indo-European languages not through reconstruction, but through naked-eye resemblances. He claims that such an approach is still warranted and useful.

As Greenberg puts it, traditional comparative linguists compare many items across a few languages (see section 1 for the reasons for doing so). Mass or multilateral comparison, as Greenberg envisions it, reverses these axes, comparing a limited number of vocabulary items across a large number of languages.

Classification for Greenberg is matching resemblances and noting a degree of resemblance. He states (1987: 5):

Is form A more like form B than it is like C? Given, for example, pan/fan/ezuk, who would hesitate? What is meant, moreover, by greater resemblance is diachronic resemblance, that is, the probability that A and B
derived by changes from a common source, as compared with C's having derived from a common source at greater remote (e.g. four/vier/cuatro) or from a different source altogether (e.g. hand/hand/mano).

This is given as his main question in LIA, which is Greenberg's latest attempt to classify languages utilizing the mass/multi-lateral comparison approach. Through his application of the method, Greenberg has culled three language families out of the Western Hemisphere. The largest of these he calls Amerind, and most of the languages in the Americas are assigned to this family.

To be scored by Greenberg, potential cognates must have resemblance in both form and meaning. Furthermore, such resemblance can occur in both root morphemes as well as affix morphemes.

Greenberg attempts to answer the question concerning what exactly constitutes resemblance in form and meaning. He presents a table of seven words and scores several as indicative of common origin (Greenberg 1955:109):

<table>
<thead>
<tr>
<th>ONE</th>
<th>TWO</th>
<th>THREE</th>
<th>HAND</th>
<th>EYE</th>
<th>EAR</th>
<th>MOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>sang</td>
<td>su</td>
<td>soti</td>
<td>yung</td>
<td>sing</td>
<td>--</td>
</tr>
<tr>
<td>2.</td>
<td>wate</td>
<td>iba</td>
<td>tati</td>
<td>ju(le)</td>
<td>no(do)</td>
<td>to(go)</td>
</tr>
<tr>
<td>3.</td>
<td>lori</td>
<td>cu</td>
<td>agozo</td>
<td>daho</td>
<td>samo</td>
<td>sumo</td>
</tr>
<tr>
<td>4.</td>
<td>ili</td>
<td>iwa</td>
<td>ita</td>
<td>ilo</td>
<td>ewu</td>
<td>212</td>
</tr>
<tr>
<td>5.</td>
<td>mwe</td>
<td>bali</td>
<td>tato</td>
<td>(le)-iso</td>
<td>(ku)-toi</td>
<td>(mu)-nywa</td>
</tr>
<tr>
<td>6.</td>
<td>illo</td>
<td>ndi</td>
<td>yasko</td>
<td>kela</td>
<td>sim</td>
<td>sumo</td>
</tr>
<tr>
<td>7.</td>
<td>kiet</td>
<td>iba</td>
<td>ita</td>
<td>eke</td>
<td>enyin</td>
<td>utong</td>
</tr>
<tr>
<td>8.</td>
<td>lakoi</td>
<td>swa</td>
<td>we</td>
<td>tahä</td>
<td>i</td>
<td>kebbe</td>
</tr>
</tbody>
</table>

gether as similar those forms which are underlined in the same style. It must be noted that I would switch the affiliations of the words for ‘one’ and ‘two’ in Alhō. I believe there is a typographical error in the Greenberg original.

Nonetheless, Greenberg still fails to explain what criteria he uses, leaving the questions unanswered. He also states that the addition of more languages to the comparison only solidifies the results.

Greenberg’s call for more languages in this case implies that mass comparison is not necessarily so massive. It appears that in Greenberg’s perspective mass comparison is simply comparison of items in anywhere from 3 to n languages, just not in pairs.

The question of scoring is indeed a tricky one, when considering Greenberg’s work. Semantic as well as formal problems have surfaced. To illustrate the basic problem, Matisoff 1990 proposed a thought experiment where the evidence is not as conclusive as Greenberg proposes (e.g., English/German/Spanish: hand/Hand/mano), but more subtle, offering ér/erku/duo ‘two’ in Mandarin Chinese, Armenian, and Latin, respectively. Which of the three does one group together? The first two have a nice correspondence in er, but the last two have a C+u correspondence. The historical situation, however, leads one to choose the latter two (for a fuller discussion of the correspondence Armenian erk-/Latin du-, cf. section 1.0.4). Mass comparison, Matisoff concludes, has no mechanism for dealing with such a problem, although it represents exactly the type of similarity that Greenberg seems to be proposing as THE similarity.
3.1 Problems with Greenberg's Approach

Greenberg's assumption that Jones (section 1) made a connection between Latin, Greek, and Sanskrit on the basis of a few words and grammatical forms is not borne out by the facts. Jones had a thorough training in both Latin and Greek, and through his legal dealings in India, had come to know a great deal about Sanskrit. Thus, Jones was making his decision based on the comparison of three COMPLETE languages, not mere words or a few small points of grammar.

It must be said that Greenberg also attempts to bring "grammatical evidence" into his method, both in Greenberg 1955 and Greenberg 1987. For example, Greenberg correlates two first-person pronouns \( n \) and \( i \), as occurring almost universally in Amerind.

However, the longest element he compares is a mere two segments long. The possibility that a resemblance of one or two segments is due to chance is incalculable. This holds true for grammatical elements as well as for free morphemes. See section 1.2.1 for a fuller discussion of the relationship between chance and short morphemes.

Fodor 1982 discusses some of the problems of scoring resemblances in Greenberg 1955. To begin with, he reviews Greenberg 1970 and 1957. In Greenberg 1957 ("Genetic Relationship among Languages"), Greenberg sets 8% resemblance as the threshold of the
relationship between two languages. Fodor states he cannot determine how Greenberg has arrived at this number, and wonders whether it is applicable to the Swadesh 200-word list or a list of many thousand items (cf. § 1.0.3 on basic vocabulary).

A major criticism, especially by Fodor, of Greenberg’s African classificatory work was that it does not make use enough of sound correspondences, sound laws, and at least the initial stages of reconstruction. Two degrees of confidence (cf. section 1.6) are in conflict here: Greenberg is willing to score as resemblances those items which are visually similar in form and meaning, while Fodor insists on more strict criteria upon which to make these judgments.

3.1.1 FORMAL AND SEMANTIC DIFFICULTIES

Also in response to Greenberg 1955, Fodor takes issue with the lack of specificity in scoring formal resemblances. He illustrates this with, among others, the following example taken from Greenberg (1955: 114): Momvu kedru ‘tongue’ and Garko jálde ‘tongue’. There are no compelling recurrent phonological reasons for placing these two forms together, aside from the fact that both mean ‘tongue’.

I have found additional formal problems in Greenberg 1987 which the literature has not addressed. Consider a list of alleged long-term correspondences proposed by Greenberg:
TO CLOSE\(^3\) (p. 205)

<table>
<thead>
<tr>
<th>Macro-Ge</th>
<th>Ge: Krenje <em>kapi</em> ‘bolt’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Puri: Coronado <em>kapo-em</em> ‘close’</td>
</tr>
<tr>
<td>Macro-Panoan</td>
<td>Lengua: Mascoy <em>kjab-</em> ‘cover’</td>
</tr>
<tr>
<td></td>
<td>Panoan: Shipibo <em>kepu</em> ‘close’</td>
</tr>
<tr>
<td>Chibchan-Paezan</td>
<td>Cuna <em>akapa</em> ‘close one’s eyes’</td>
</tr>
<tr>
<td>Hokan</td>
<td>Tonkawa <em>kapa</em> ‘shut’</td>
</tr>
<tr>
<td>Almosan-Keresiuan</td>
<td>Proto-Central-Algonquian <em>kep</em> ‘to close’</td>
</tr>
<tr>
<td></td>
<td>Siouan: Catawba <em>kopa</em> ‘to close’</td>
</tr>
</tbody>
</table>

All the items cited by Greenberg are indeed very similar in form. My argument is that they are actually too similar, if one is to accept the many thousand years’ time depth proposed.

By way of contrast, consider an Indo-European word taken from the list of basic vocabulary: English *horn*. In the closely related Germanic languages, the word is very similar: German *Horn*, Dutch *hoorn*, Swedish *horn*. However, as one observes related words in the other branches, the forms are more divergent: Latin *cornus*, Greek *kéras*, and especially Hindi *sík*, which bears hardly any discernable similarities to the other words.

The Hindi and English forms can be explained by sound shifts, which have acted upon the original Indo-European root to produce the forms we have today. The time depth necessary to reconcile nearly all the languages in the Americas is surely as great as that separating Proto-Indo-European from today, yet the expected ef-

---

\(^3\) Perhaps (on formal grounds) also ‘to cap’!
fects of sound change do not seem to materialize in Greenberg's data.

While it is possible that sound change operates very slowly on these languages (cf. the slow rate of change in the Turkic languages), the existence of a reconstructed protoform in the data suggest a reasonable time depth, which should produce more sound change than is exhibited in the data.

As we attain a deeper level of linguistic history, Chafe claims, the quality of the data often is diminished, due to the limited scope of vocabulary that can be reconstructed as well as increasingly shorter roots. Mass comparison is supposedly, according to Greenberg, "so powerful that it will give reliable results even with the poorest of materials" (1987: 29).

Yet Chafe demonstrates definite failures of this "powerful" method, where poor materials have yielded inaccurate scorings. He cites Greenberg's comparison of Seneca *skə 'one' and Caddo *wist'i 'one' (P. 175 in G.; p. 653 in C.). Chafe points out that the Seneca word is actually *skə:t, and that the -t element missing from Greenberg's citation is the historical root, meaning 'stand', while ka- is a neuter prefix and s- is a repetitive prefix. The actual Caddo word is *wist's'i?, but its etymology is unclear. If Seneca and Caddo are related, the evidence is not furnished by these examples.

As further illustration for his case regarding the power of Greenberg's method, Chafe inspects Greenberg's entry for 'full': Proto-Central Algonquian *moːt'ki 'fill', Proto-Salishan *məq' 'eat
one’s fill’ and Seneca -nōke ‘abound’ (p. 171 in G.; 653 in C.). Greenberg follows the Seneca entry with the notation, (n < *m), which he attributes to Chafe. Chafe denies that he has claimed such an origin for Seneca n, and says that no one, to his knowledge, has even suggested it. Thus taking away part of the support Greenberg has claimed for this relationship, Chafe then goes on to call into question the validity of the forms that are compared, and, to an extent, the level on which they are compared.

Chafe offers *-nakre- as the Proto-Northern-Iroquoian form of Seneca -nōke. If Greenberg is not to violate his own stated principle (cf. 3.2.3) of not comparing isolated languages with wider groupings, he should compare Proto-Northern-Iroquoian, rather than Seneca, with Proto-Central-Algonquian and Proto-Salishan. Chafe also takes issue with the meaning that Greenberg gives for -nōke, stating that it is a misinterpretation of the gloss given in Chafe (1967): “be present in abundance, dwell together in a group.” He says the most likely original meaning was ‘dwell together’. Chafe states:

Comparing Proto-Northern-Iroquoian *-nakre- ‘dwell together’ with Proto-Algonquian *mo:t'ki ‘fill’ or Proto-Salishan *məq ‘eat one’s fill’, assuming that these last two are valid reconstructions, hardly provides cause for rejoicing over the genetic unity of “Almosan-Keresiouan.” (653)

Another central concept to the question of scoring resemblances is the notion, provided by Greenberg 1957, of semantic plausibility. Campbell points out that Greenberg has left too wide a range in semantics, quoting such examples as ‘excrement/night/grass’ and
'child/copulate/son/girl/ boy/tender/bear/small'. Campbell (1989: 600) notes that "...any necessary but undocumented assumption of semantic shift increases the possibility that chance, rather than common ancestry, explains a phonological similarity." The reader is reminded of Greenberg's quotation in section 1.5 (above), "...the greater the methodological latitude permitted, the less plausible is each individual comparison."

Campbell 1988 remarks that in several instances, Greenberg has compared words on the basis of similarities in form and (extended) meaning. The known details, however, do not support Greenberg's assumptions in all cases. Greenberg correlated Kekchi č'oc'el 'liver' with Quiché kusč 'heart'. Campbell states, however, that the Kekchi and Quiché words are actually from two separate etyma (Proto-Mayan *č'a 'earth', and Proto-Quichean *k'usč 'heart, chest', respectively). Thus, the resemblance between Quiché and Kekchi in this case is due not to genetic relationship, but to chance.

---


3.2 Recurrent Problems

The problems encountered in Greenberg's method are the same ones that comparative linguists have wrestled with since the inception of the field. Most of the criticisms of mass comparison (Fodor 1982, Chafe 1987, Campbell 1988, Matisoff 1990) have pointed to chance similarities and borrowings as problems, and I shall discuss these, following the general outline used in section 1. In addition, also as in section 1, I will consider the problems of onomatopoeia and nursery language, discussion of which is absent from most of the critical literature regarding Greenberg.

3.2.1 Chance Similarities

Chance as a source of resemblance was noted by Chafe 1987. As an example, he offers Tuscarora -tuʔθ-, which looks and sounds very similar to English tooth. However, as we go back in time, the resemblance is less, as the Tuscarora word comes from an earlier *-noʔtʃ-, and the English one comes from PIE *d̪ont-. As more languages are added to the comparison, the possibility that chance has produced similarities is also increased. Chafe finds no methods for coping with chance similarities, or for that matter, borrowing. He writes (1987: 652), "One needs, of course, criteria that will go beyond just a feeling that since two words look alike, they must have a common origin."
3.2.2 BORROWING

Matisoff 1990 discusses Greenberg’s reluctance to believe in the ability of lexical and grammatical traits to diffuse across genetic boundaries. Matisoff points out that there are instances where even large amounts of basic vocabulary have been borrowed. As an example, he offers Bai (Minjia), whose exact location within the Sino-Tibetan family is not known, due to the fact that as much as 75% of its vocabulary has been borrowed from a fairly closely related neighbor -- Chinese. His discussion of areal traits centers around Southeast Asia, where many Austronesian and Austroasiatic languages have a number of deep grammatical elements in common (e.g., a causative morpheme in -p-), yet they share almost no core vocabulary. Matisoff therefore concludes that the resemblances between the Austronesian and Austroasiatic languages are due to areal effect, not to genetic unity. The reader is reminded of Greenberg's (1957) Thai, Sino-Tibetan, and Kadai/Malayo-Polynesian example presented in section 1.1.

Matisoff’s conclusion is that Greenberg’s method is unable to detect chance resemblances. Matisoff points out that Greenberg is mistaken in believing that borrowing can occur only between (daughter) languages; it can occur between protolanguages as well (cf. § 1.1 above).
3.2.3 **Greenberg's Response**

Greenberg states that he is aware of the problem of borrowing and/or chance similarities. The way in which his method deals with these basic problems is a reversal of the traditional means, much as his method is a reversal of the traditional comparative method. Instead of insisting on a large number of correspondences between two languages, Greenberg's method considers the "... importance of resemblances recurrent in a large number of languages ...". The question then becomes: in light of the possibility of chance similarities and borrowings between protolanguages (cf. sections 1.1.1 and 1.2.2), is a single correspondence found in 30 languages more telling than 30 correspondences found in each of two languages?

Greenberg seems to scoff at chance and borrowing occurring between languages and families (or protolanguages). In response to Campbell, comparing Finnish to Amerind, Greenberg states 'I would never compare Finnish in isolation. If Finno-Ugric and the larger Uralic group to which it belongs were not already recognized, I would have discovered them' (Greenberg 1989: 111).

Yet, what would his response be if Finnish had no close relatives, and he was unaware of this at the outset? If that were the case, Finnish would function as a type of protolanguage, and the chance resemblances between Finnish and Amerind would be of the same nature as between Indo-European and Afro-Asiatic, assuming that
Indo-European or Afro-Asiatic had only one surviving member. This actual state of affairs would thus nullify his counter-argument.

3.1.3 ONOMATOPOEIA AND NURSERY LANGUAGE

Greenberg uses as evidence items which are likely to be due either to onomatopoeia or nursery language. To illustrate the point, I present just a few examples from the Amerind dictionary in Greenberg 1987:

42 BLOW (v.)

Macro-Ge BLOW
Mashakali: Manuci paepi ‘wind’.
Puri: Coronado pui

Macro-Panoan BLOW
Lule phu. Tacanan: Proto-Tacanan *peī

Central Amerind BLOW
Uto-Aztecan *puti.

Kiowa-Tanoan: Kiowa pho(-le).

Penutian BLOW
Oregon: Alsea pu.

Maidu pi

If one takes into account the various words related to ‘blowing’ or ‘wind’ in different languages (cf. English puff), the conclusion is that the words cited under ‘to blow’ may be onomatopoeic and not indicative of genetic relationship (cf. § 4.8.3 below).
14 AUNT

Macro-Panoan AUNT
Mataco: Macca *nana* ‘mother’.
Tacanan: Proto-Tacanan *nene

Equatorial AUNT
Kariri *aṇa*

Macro-Tucanoan AUNT
Mamainde *nat* ‘mother’. Maku *nō*
‘mother’. Cataquina *inai*

Central Amerind MOTHER
Uto-Aztecan *na*

Almosan-Keresiuan MOTHER
Caddoan: Caddo *ʔi-naʔ* ‘my mother’.
Iroquoian: Huron *anan* ‘aunt’

Because of the predominance of nasals and the vowel /a/ (cf. section 1.4 for further discussion), this example may well be simple nursery language instead of being indicative of genetic relationship. This set of words should therefore not be used as the basis for establishing genetic relationship.

These are but two of numerous potential problems with onomatopoeia and/or nursery language. Care needs to be exercised in judging which words do indeed show probability of common origin, as was recognized by comparative linguists when devising the comparative method.

4.0 THE CURRENT STUDY

In the course of the controversy regarding mass comparison, there has been (to my knowledge) no study undertaken to test Greenberg’s methodology against languages whose history is known.
For such a test, I have chosen to compare English, Hindi, and Finnish. I have chosen these three languages for several reasons. The first was convenience. In limiting the languages, I am better able to explain phenomena, yet I am not doing solely pairwise comparisons, as has been criticized by Greenberg.

English, Finnish, and Hindi are not close neighbors geographically, thus limiting the possibility of areal influences among them. There is, however, an interesting distribution of genetic and areal factors. Finnish and English are in the same basic cultural sphere (i.e., Europe), but stem from different linguistic stocks. Hindi and English, on the other hand, are members of the same linguistic stock, but are in different cultural areas.

As Hock (personal discussion) has pointed out, areal factors seem to play a role in lexical similarities. Thus, while the list of vocabulary items in Greenberg 1957 makes it fairly easy to establish the (degree of) linguistic relationship in Europe, the non-European Indo-European languages are significantly more difficult. Even in onomatopoeia, areal considerations seem to play a role. Baltic and Slavic on the Eastern edge of Europe use onomatopoeia freely in all kinds of literary genres. English uses onomatopoeia quite rarely in nearly all forms of written discourse (although a fair amount of use is found in spoken colloquial English). German takes something of a middle position.

Greenberg states many times in *Language in the Americas* that the main task of historical linguistics is to show whether certain lan-
languages are closer to each other than to any other languages in establishing a valid genetic unity. Mass comparison, as applied to the three languages in this study, could yield any one of several possible scenarios:

The method could show that all three are more or less equidistant, that Hindi and English are closer to each other than to Finnish, that Hindi and Finnish are closer to each other than to English, or that English and Finnish are closer to each other than to Hindi. Which one will it show, and will this scenario be conclusive in comparison to the other possibilities?

In short, I wish to explore a fundamental question in this study. To what extent do the similarities reflected by the application of mass comparison tell the whole story regarding the relation of the languages in question? In other words, do these similarities reflect the actual relationship of the languages? Will the method of mass comparison correctly show that English and Hindi are more closely related to each other than either are to Finnish?

In scoring items as cognates, I have been guided by the principles of looking for degrees of phonetic and semantic similarity that are, I believe, substantially the same as Greenberg’s. In the following presentation, I list first the words in each language pair/triplet found to be similar. I then list the consonant and vowel correspondences in the pair/triplet and present an analysis via the comparative method. I note the correspondences in the various environments in which they occur, arranged phonologically with
reference to the first language listed, in order to facilitate analysis.

Finally, I present the word pairs/triplets in light of known etymologies, (Collinder 1955 and 1957, Drosdowski 1963, Mayrhofer 1956-1980, Partridge 1983, Skeat 1911, Toivonen 1955-1969, Watkins 1985. Hock provided me with Sanskrit etyma of Hindi words.) as a final check on the method's ability to predict genetically related items. If I have found a word to be borrowed from an outside source (i.e. not one of the three languages in the study), I have decided to reject the item as indicative of genetic relationship, because the borrowing does not reflect an inherited root. However, there is the case of Hindi borrowing a word from Sanskrit, its mother language. If this occurs, I score the word as both a borrowing and inherited, one at a time.

4.1 THE COMPOSITION OF INDO-EUROPEAN AND URALIC

In order to show the generally accepted classification of these languages, as well as to acquaint the reader with the other closely related languages, I wish to review the composition of the Indo-European and Uralic families.

Indo-European consists of eleven major branches (given here in alphabetical order): Albanian, Anatolian, Armenian, Baltic, Celtic, Germanic, Hellenic, Indo-Iranian, Romance (Italic), Slavic, and Tocharian.

Albanian

Anatolian: Hittite
Armenian

Baltic: Latvian, Lithuanian, and Old Prussian

Celtic: Brythonic: Welsh, Cornish, Breton.
Goidelic: Irish, Manx, Gaelic

Germanic: East Germanic: Gothic
West Germanic: English, Frisian, Dutch, Afrikaans,
Yiddish, and German, both High and Low
North Germanic: Swedish, Danish, Norwegian
(Bokmål and Nynorsk), Icelandic, Faroese

Hellenic: Modern Greek, ancient dialects of Greek (Attic, Aeolic,
Mycenaean, Ionic)

Indo-Iranian: Indic: Sanskrit, Hindi, Bengali, Kashmiri, Urdu, etc.
Iranian: Modern Persian, Kurdish, Pashto, Avestan

Italic: Latin, Spanish, Portuguese, Catalan, Provençal, French,
Italian, Romansch, Romanian

Slavic: Eastern Slavic: Russian, Byelorussian, Ukrainian
Western Slavic: Polish, Czech, Slovak
Southern Slavic: Bulgarian, Serbo-Croatian, Slovenian

Tocharian: Tocharian A, Tocharian B

Uralic can be divided into two groups. These are Fenno-Ugric and
Samoyed. Fenno-Ugric, in turn, consists of six groups: Baltic Fennic,
Lappish, Mordvin, Cheremis, Permian, and Ugric.

Baltic Fennic: Finnish, Estonian (containing two major dialects),
Karelian, Olonets, Veps, Lude, Vote, Livonian

Lappish
Mordvin: Ezra, Moksha
Cheremis
Permian: Votyak, Ziryene
Ugric consists of three languages, divided into two groups.
Ob-Ugric: Vogul, Ostyak
Hungarian

The Samoyed languages are divided into two groups: Northern and Southern. Yurak, Tavgi, and Yenisei Samoyed make up Northern Samoyed, while Selkup and Sayan Samoyed (Kamassian) comprise Southern Samoyed.

4.3 SELECTING AFFINITIES

The putative correspondences were established by scanning the list of the English glosses of 84 core concepts taken from Greenberg 1987, and comparing these to Hindi and Finnish equivalents. Selection was made on the basis of rough similarities encountered while visually scanning the list. To the best of my ability, I have tried not to let prior knowledge of cognates affect my judgment.

There were, to be sure, difficulties in selecting the groups (pairs or triplets) that should be regarded as 'bespeaking a common origin'. The question that arises is, how close do words have to be in order to be scored as similar? Are English saliva and Finnish sylki sufficiently close, and, if they are, are Hindi pissū 'flea' and Finnish kirppu 'flea'? I have found nowhere in Greenberg's work explicitly stated criteria by which to make this judgment.
Similarly, consider the case of English *clean* and German *klein* ‘small’. These two words are known to be historically related, and the phonological similarity can still be seen, yet the meanings have diverged significantly from the original meaning of ‘pure’ or ‘shiny’. How would this pair be scored by Greenberg?

Most likely, he would accept such a pair, positing the intermedi-ate steps in the shift (i.e. that small items are more likely to be kept clean, and so on), as he has similarly done in *L/A*, with supporters coming to his defense (Shevoroshkin, 1990:24):

Some scholars criticized Greenberg for the seemingly arbitrary comparison of words that sound similar in native tongues but have largely different meanings, such as the terms for feces, night and grass. But Greenberg’s analysis was far from capricious. All three terms are derived from the original Amerind word for dirt; with time the word came to mean dirty, as in excrement, and later simply dark. The term subsequently took on specific color connotations, such as black (a reference to night), blue and in some languages green -- the color of grass.

While it is possible to see the connections being drawn here, the reader is referred to the quote by Greenberg from *Essays in Linguistics* cited above (§ 1.5.1).

In my study questions of this nature have arisen. For example, in addition to English *carry* and Finnish *kärsiä* ‘carry’, Hindi has a word *kar*, meaning ‘hand’. Is this to be considered or disregarded? While I have chosen to accept it, there is no methodological criterion on which to base such a decision — or to reject it.
Depending on the amount of latitude granted in semantic matters, there are two possible consequences. Either too many words are considered to be related, or not enough.

The concepts 'carry' and 'hand' are probably in closer semantic fields than are 'feces' and 'night'. Yet we do not have any other further justification for *carry*, *kärsiä*, and *kar* other than the outward similarity in form and the extended relation in meaning. We thus could be grouping together more than we are justifiably able to.

On the other hand, if we simply reject this latitude, we might group together less than we should. As was cited earlier, English *clean* and German *klein* are related, but this relationship would be missed if we did not allow for some semantic shift. My approach has been to allow for a reasonable range of meaning, in order to give Greenberg's new method the benefit of the doubt.

4.4 PHONOLOGICAL SYSTEMS OF FINNISH AND HINDI

In order to understand the possible correspondences proposed among the languages in question, it is necessary to understand the phonological systems involved. As the reader is probably familiar with English, I will present here only the phonological systems of Finnish and Hindi.

Finnish:

<table>
<thead>
<tr>
<th>type</th>
<th>labial</th>
<th>apical</th>
<th>velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced</td>
<td>(b)</td>
<td></td>
<td>(g)</td>
</tr>
<tr>
<td>stops</td>
<td></td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>voiceless</td>
<td>p</td>
<td>t</td>
<td>k</td>
</tr>
</tbody>
</table>
voiced \( v \)
fricatives
voiceless (f) \( h \)
nasals \( m \) \( n \) \( \eta \)
lateral \( l \)
trill \( r \)
sibilant \( s \)
semivowel: \( j \)
vowels:
front
center
back
unround round
high \( i \) \( y \) \( u \)
mid \( e \) \( ö \) \( o \)
low \( å \) \( a \)

Additionally, there are regular consonant alternations, such as:

\[ pp: p \quad \text{appi 'father-in-law': gen. ape-n} \]
\[ mp: m \quad \text{ampu-a 'to shoot': ammu-n 'I shoot'} \]
\[ p: v \quad \text{kipu 'pain': gen. kivu-n} \]
\[ tt: t \quad \text{kontti 'knapsack': gen. konti-n} \]
\[ nt: nn \quad \text{anta-a 'to give': anna-n 'I give'} \]
\[ rt: rr \quad \text{parta 'beard': gen. parra-n} \]
\[ lt: ll \quad \text{ilta 'evening': gen. illa-n} \]
\[ t: d \quad \text{pata 'dike': gen. pada-n} \]
\[ lk: l \quad \text{jalka 'foot': gen. jala-n} \]
\[ lk: lj \quad \text{jalki 'footstep': gen. jälje-n} \]
\[ rk: r \quad \text{märkä 'wet': gen. märä-n} \]
\[ rk: rj \quad \text{särke-ä 'to break': särje-n 'I break'} \]
\[ hk: hj \quad \text{pohkee-n gen. of pohje 'calf'} \]
\[ ηk: ηη \quad \text{lanka /länka/ 'thread': gen. langa-n /länna-n/} \]
\[ Vik: Vj \quad \text{poika 'boy': gen. poja-n} \]
\[ k: v \quad \text{puku 'dress': gen. puvun} \]
4.5 ENGLISH-FINNISH

The correspondences between English and Finnish yield results of 7/84 or 8.3%. The list includes two obvious borrowings, *penis/penis* and *person/persoon*, but they match exactly Greenberg's criteria of correspondence in form and meaning. Excluding these borrowings, the numbers are 5/84 or 6%, below Greenberg’s 8% from 1966, but well above the 2.2 percent proposed in 1987.

1. clean (v.)  kiillottaa
2. fly          lentää
3. hair         hius
4. penis        penis
5. person  
6. powder  
7. puny  

4.5.1 **CONSONANT CORRESPONDENCES BETWEEN ENGLISH AND FINNISH**

Initial: k-k (1), h-h (3), p-p (4,5,6,7), fl-l (2)

Medial: l-ll (1), r-r (5), s-s (5), n-n (4,7), d-l (6), ø-nt (2)

Final in English-Medial in Finnish: n-tt (1), n-n (5)

Final: r-s (3), s-s (4)

In the data relating English with Finnish, some rather interesting correspondences arise. The most recurrent correspondences are *p* and *p̥*, which occurs four times, always in initial position, as well as *n* and *n̥*, which occurs three times, twice in medial position and once in final position in English and medial in Finnish. Both the *n* and *p* correspondences occur twice in obvious borrowings. The remaining initial correspondences are all equational, with the exception of *fl-l*, problems with which are alleviated by the fact that Finnish does not have initial clusters.

The medial correspondences are relatively straightforward, with the only exceptions being *d-l* and *ø-nt*. While there are a fair number of consonant variations in Finnish, these two correspondences are not among the gradation examples. Much the same can be said of the other correspondences. The one feature that is common to the non-equational correspondences is that they are apical. There are also two instances of a single consonant in English corresponding to a geminate in Finnish.
4.5.2 Vowel Correspondences between English and Finnish

First syllable: i-o (1), i-e (4), e-iu (3), æ-e (5), au-ö (6), (j)u-ie (7), ai-e (2), zero-ii (1)
Second syllable: ø-i (4), ø-oo (5), i-i (7), ø-ü (6), zero-AA (1)

There are no recurrent vowel correspondences between English and Finnish.

4.5.3 Etymological Considerations

Let us take a closer look at these pairs, considering the etymological evidence. Results are arranged according to borrowings, chance similarities, or actual genetic relationship.

Borrowings:

**PENIS: PENIS** This anatomical word, both in English and Finnish, comes from Latin *pēnis* ‘tail, male member’.

**PERSON: PERSOONA** Although the exact origin of this word is unclear, it is simply a borrowing, in both languages, from Latin *persona*.

Chance:

**CLEAN (v.): KIILLOTTAA** The English word stems from the Germanic *klainī- ‘clean, pure’, which is further related to Indo-European *gel- ‘bright’. The Finnish word is widely attested in Fenno-Ugric (cf. Karelian *kiildä*, Estonian *kiilas*, Livonian *kiłaz*). There is no evidence that the two words are related.
FLY: LENTÄÄ This pair was selected on the basis of the Finnish lack of initial clusters, along with a liberal view of possible correspondences. The English word goes back to an Indo-European root *pleuk-, which probably had the meaning ‘to move quickly’ (cf. Lithuanian plaũkti ‘to swim’). Cognates of the Finnish word appear to be limited to the Baltic Fennic area (cf. Karelian lendiä, Lude lendiä, Veps leta, Estonian lennata), making it unlikely that the word has been inherited from Proto-Uralic. In addition, the phonetic differences between the Indo-European and Baltic Fennic forms, makes genetic relationship doubtful.

HAIR: HIUS English hair (Old English hær) reflects the Germanic root *haz-. The Germanic root in turn reflects Indo-European *k’er[s] ‘to be raw, to be shaggy’ (cf. Lithuanian šerųš ‘brush’, Russian šerši ‘wool’). The Finnish word has outside relatives only in the Baltic Fennic subgroup of Uralic (cf. Karelian hivuš, Lude hivus, Veps hibus, Estonian juus, Livonian ibûks). The limited distribution raises doubts on its being inherited from Proto-Uralic. Furthermore, lack of similarity with the Indo-European root makes genetic relationship unlikely.

POWDER: PÔLY The English word is most likely borrowed from Latin puluerem, which in turn is derived from puluis ‘dust’. A Fenno-Ugric root is reflected in Finnish (cf. Karelian pölä, Lude pölä, Votyak pöllä, Veps pölä). Because the English word has been borrowed from Latin, genetic relationship cannot be established on the evidence of this word.
The English word comes from Middle French prisme ‘small’, which in turn is derived from Latin post nātus ‘born late; newly born’. The cognates of the Finnish word are widely attested in the Fenno-Ugric languages (cf. Karelian pieni, Lude pien, Veps peń, Votyak řeni). Genetic relationship is highly unlikely, because the English word was borrowed. Furthermore, the French word derives from a Latin phrase, not a single root. The similarities pie- : pu- (< Latin post ‘after, later’) and -ni : -ny (< Latin nātus) thus are entirely accidental.

**Genetic Relationship:**

None of the similarities reflected in this section is due to genetic relationship. Furthermore, none of the roots listed in this section has been proposed by scholars like Collinder (1957) as possible connections between Uralic and Indo-European.

### 4.6 HINDI-FINNISH

The similarities between Hindi and Finnish amounted to 16/84, or 19.1%. The words did not appear to be concentrated in any one semantic field.

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Finnish</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>bhāy</td>
<td>pelko</td>
<td>“fear, anxiety”</td>
</tr>
<tr>
<td>dhūdhla</td>
<td>tietämätön</td>
<td>“dark, black”</td>
</tr>
<tr>
<td>galā</td>
<td>kaulanetuosa</td>
<td>“throat”</td>
</tr>
<tr>
<td>hāl mē</td>
<td>eilen</td>
<td>“yesterday”</td>
</tr>
<tr>
<td>(h)ōth</td>
<td>huuli</td>
<td>“lip”</td>
</tr>
<tr>
<td>ūr</td>
<td>juuri</td>
<td>“root”</td>
</tr>
<tr>
<td>kākī (father’s brother’s wife)</td>
<td>tāti</td>
<td>“aunt”</td>
</tr>
</tbody>
</table>
8. kān  korva  "ear"
9. kukkur  koira  "dog, hound"
10. nāk  nokka  "nose"
11. nigalnā  niellä  "swallow (v.)"
12. pata lagānā  pyytää  "seek, look"
13. stan  sydän  "breast"
14. shishn  siitin  "penis"
15. sīg  sarvi  "horn"
16. tiţnā (v.i.)  taittaa  "break (v.)"

4.6.1 CONSONANT CORRESPONDENCES BETWEEN HINDI AND FINNISH

Initial: bh-p (1), dh-t (2), g-k (3), p-p (12), t-t (16), k-t (7), k-k (8,9), st-s(y)d (13), sh-s (14), s-s (15), h-(e) (4), (h)-h (5), j-j (6), n-n (10,11)

Medial: dh-t (2), t-t (12), t-t (16), th-l (5), k-t (7), k-ø (11), g-ø (11), n-ø (16), l-m (2), l-l (3,11)

Final: th-l (5), k-k(V) (10), g-r (15), n-n (13,14) n-r (8), r-r(V) (9), r-r (6), l-l (4)

If one compares kān and sīg with korva and sarvi, and if we assume that the nasalization on the vowel in sīg carries over to the /g/, then an interesting correspondence obtains, namely N-rv. Alternately, it is possible to link the g in sīg with a k in Finnish, which has regular gradation of k with v.

4.6.2 VOWEL CORRESPONDENCES BETWEEN HINDI AND FINNISH

First syllable: ai-e (1), ū-ie (2), a-au (3), ā-ei (4), ā-uu (5), a-uu (6), ā-ā (7), ā-o (8,10), u-oi (9), i-i (11), a-yy (12), a-ā (13), i-ii (14), ī-a (15), ū-ai (16)
Second syllable: ā-ā (2), ā-a (3), ī (7), u-zero (9), a-e (11), ā-āā (12), ā-aa (16)

Third syllable: ā-ā (11)

Zero in Hindi-Vowel in Finnish: zero-o (1), zero-ā (2),
zero-i (5,6,14,15), zero-a (8,9,10), zero-y (12)

Several correspondences occur multiply. The most recurrent is
between lack of a vowel in Hindi and /i/ in Finnish, which occurs
four times. However, lack of a vowel in Hindi also corresponds to
many other vowels in Finnish (/a/ being the second most
prevalent). The other multiple occurrences are /ā/ in Hindi with
/o/ in Finnish (in the first syllable only), and /ā/ in Hindi with /ā/
in Finnish (in both first and third syllables). The other
correspondences are single instances.

4.6.3 ETYMOLOGICAL CONSIDERATIONS

Let us take a closer look at these pairs, considering the etymological
evidence. Results are arranged according to borrowings, chance
similarities, onomatopoeia, nursery language, or actual genetic relation-
ship.

Borrowing:

GALĀ: KAULANETUOSA The Hindi word reflects Indo-European
*gel-*ger- (cf. German Kehle ‘throat’, Latin gula ‘esophagus’).
The Finnish word, along with related Karelian (kagla) and Estonian
(kael), appear to have been borrowed from Baltic (cf. Latvian kakls, Lithuanian kāklas), according to Collinder (1955).

**ŚĪG: SARVI** The Hindi word is from Sanskrit śīṅga, which is ultimately related to English horn (note that mass comparison would not have selected the Hindi word as related to English). While the word sarvi is widespread in Uralic (cf. Hungarian szarv, Lappish čoarve, Cheremis šur-), the source, according to Collinder, is ultimately Avestan srvā (a cognate of Sanskrit śīṅga).

**Chance:**

**BHAY: PELKO** The Hindi word stems from Sanskrit bhaya-, which, according to Mayrhofer, is related to other words outside of Indo-Iranian (cf. Lithuanian bāime, Latvian bitiēs ‘to be afraid’, Old Church Slavonic bojati sq ‘to be afraid’). The Finnish word has widespread correspondences in Uralic (cf. Hungarian fél-, Mordvin pele- and Yurak Samoyed pīrū-), and Collinder offers *pele-* as the Proto-Uralic form.

**DHŪDHLĀ: TİETĀMĀTŌN** The etymology of the Hindi word is not known exactly, but Mayrhofer suggests that there may be a Dravidian connection via the Sankrit word dhūrtam ‘rust’. He declines to speculate further. The Finnish word was unavailable in my etymological sources.

**HĂL MĔ: EIİEN** The exact etymology of the Hindi word is unclear, but the Sanskrit equivalent is hyāḥ ‘yesterday’. The Indo-European root of the Sanskrit word was *gḥ(y)es-* ‘the other day’, which bears
little resemblance to the Uralic forms. Cognates for the Finnish word are found only in Baltic Fennic (Karelian eglen, eklen, Olmets, Lude, Veps egläi, Estonian eile). The resemblance between the nasals (Hindi me, Finnish eilen) is misleading, because, according to Hock, the Hindi postposition is an innovation, probably developed from Sanskrit madhya- ‘middle’.

(H)ðTH: HUULI The source for the Hindi word was Sanskrit oṣṭhaḥ. Mayrhofer connects the Sanskrit word to Indo-European *au- (cf. Latin oṣṭium ‘mouth of river’). The Fenno-Ugric word for ‘lip’ shows little variation, except for some alternation of the initial h (cf. Karelian huuli, Lude hūli, Votyak ūli, Estonian ūl). Comparing the Fenno-Ugric words with the Indo-European root illustrates that this is clearly a chance similarity.

JÅR: JUURI The Hindi words comes from Sanskrit jata ‘matted hair’. The etymology of the Sanskrit word, according to Mayrhofer, is not satisfactorily explained. Cognates of the Finnish word (cf. Olonets juuri, Lude dūr, Votyak jūri, Mordvin jür) show variation in the initial consonant, although j predominates. It is unlikely, however, that the resemblances are the result of genetic relationship, especially taking into consideration that the -r- in the Hindi word which corresponded so well to the Finnish word actually comes from a -t-.

KÄN: KORVA The Hindi word stems from Sanskrit karṇa, which reflects an innovation in Indo-Iranian (the Indo-European root for ‘ear’ was *ōus-). The basic shape of the Finnish word is reflected
throughout Fennougric (cf. Karelian korva, Olonets korvu, Veps korv, Estonian körv). Consideration of the Indo-European form reduces significantly the possibility that the form is an inherited one.

NÄK: NOKKA The Hindi form is an innovation in Hindi, which is probably the result of irregular sound change from Sanskrit. The Sanskrit word is nāsa, which is related to English nose (cf. section 4.7 below). Cognates with the Finnish word, which conveys the meaning ‘beak’ more than ‘nose’, are widespread in Fennougric (cf. Olonets Ṉokku, Karelian Ḉokko ‘nose’, Lude ńokk, Estonian nokk, Ostyak no:x, Vogul ńčkw-). Because the /k/ in the Hindi form is recent, it is unlikely to be related to the /k/ in Fennougric, and thus calls into question the relationship of the complete words.

NIGALNÄ: NIELLÄ The Hindi form reflects Proto-Indo-European *ni-g(w)er-/ni-g(w)el-, which contains the root for ‘throat’ found above under galā. A widespread Uralic root is reflected in the Finnish word, for example, Hungarian nyel-, Cheremis nela- and Tavgi Samoyed ńaltami-. However, the most salient feature in the resemblance of the Hindi and Finnish words is the ni- element. In Indo-European, this was a prefix meaning ‘down’, whereas in Fennougric, it is part of the root for ‘throat’.

PATÄ LAGÄNÄ: PYYTÄÄ The source of the Hindi word/expression is unclear. Pyytää, which also conveys the meaning ‘to hunt’, is well attested in the Baltic Uralic area (cf. Yurak Samoyed piu-/püu- ‘to search, to hunt’, Lappish biwde- ‘to hunt, to fish’).
STAN: SYDÄN  The Hindi source is Sanskrit stánah, further related to Avestan fšānā- 'female breast' and Lithuanian spenys 'nipple'. A Uralic root, *sūdāme 'heart, chest', is reflected in Mordvin sedej 'heart', Cheremis šūm 'heart', and Hungarian szive 'heart', as well as in the Finnish word meaning 'chest'. The Indo-European root most likely had an initial cluster (cf. Avestan), whereas the Uralic root lacks this. Additionally, the Uralic root has a vowel between the alveolar fricative and stop elements, making genetic relationship unlikely.

SHISHN: SHITIN  The Hindi word stems from Sanskrit šisṇám 'tail, penis'. Mayrhofer offers no further etymology of the Sanskrit word, implying that it does not stem from Indo-European. The Finnish word was unavailable in my etymological sources.

TUȚNÄ (v.i.): TAITAA  According to Mayrhofer, the source of the Hindi word is Sanskrit truṭāti 'to fall apart'. The Finnish word, also related to Finnish taima 'ridge of house' and taimi 'curvature (especially of sled runner)', is related to Lappish doaggje~doaje 'fold up, break, crack' and Cheremish taże 'to bend down'.

Onomatopoëia:

KUKKUR: KOIRA  The Hindi word has an antecedent in Sanskrit kurkura, which can be attributed to onomatopoëia (the Indo-European form is listed under English hound in section 4.7 below). The Finnish word, according to Toivonen, appears to have its origins in the Uralic designation for 'male' (cf. Hungarian here 'testicle',
Yurak Samoyed *hoora* (male) reindeer, Vogul *kaar/haar* 'male reindeer'). The fact that one of the words is onomatopoeic while the other designates 'male' in general does not lead one to the conclusion that these are related words.

**Nursery Language:**

*Kākī Tāti* Neither the Hindi nor the Finnish word was unavailable in my etymological sources. Both are, however, clearly nursery words, and do not represent genetic relationship.

**Genetic Relationship:**

There were no words reflecting genetic relationship in this list, and none has been considered by the Collinder as indicative of connections between Uralic and Indo-European.

### 4.7 ENGLISH-HINDI

A comparison of the 84 concepts reveals correspondences between English and Hindi in an impressive 23.8% of the words, or 20/84.

1. boil (v.) uhalnā
2. brother bhāī
3. chant gānā
4. chest (body) ċhātī,
5. chew čabānā
6. dust dhūl(i)
7. far dūr
8. fear ḍar,
9. grass ghās
10. hound  kuttā
11. little  laghu
12. mouse  mūsikā (house), mūs (field)
13. near  nazdīk (i)
14. nose  nāsā
15. red  rakt
16. shoulder  skandh
17. thin  tanu (things)
18. two  dō
19. wish (v.)  āshā karnā
20. young javān

4.7.1 CONSONANT CORRESPONDENCES BETWEEN ENGLISH AND HINDI

Initial: θ-t (17), t-d (18), b-bh (2), b-(u)b (1), d-dh (6), g(r)-g (9), f-d (7), f-ḍ (8), s-sk (16), h-k (10), ċ-g (3), ċ-čh (4), ċ-č (5), l-l (11), r-r (15), m-m (12), n-n (13,14), w-ñ (19), y-ṛ (20)

Medial: ḍ-ð (2), t-g (1), l-n (16), d-dh (16)

Final: t-ṭ (3), t-t(i) (4), l-l (1), nt-n (3), st-t (4), st-l (6), r-r (7,8), r-z (13), s-s (9,12), nd-t (10), l-ñ (11), z-s (14), d-(k)t (15), n-n (17), ś-s (19), ṇ-n (20)

Several correspondences occur multiply: in initial position /n/ and /n/ (twice), in final position (in English) /r/ and /r/ (twice: note also /r/ and /z/, which is not all that unusual (cf. Germanic rhotacism), and /s/ and /s/ (twice: note also /z/ and /s/ occurring once).

4.7.2 VOWEL CORRESPONDENCES BETWEEN ENGLISH AND HINDI

First Syllable: oi-a (1), θ-ā (2), θ-ū (6), ə-a (20), ə-ā (3,9), ɛ-ā (4), u-a (5), u-ō (18), a-ū (7), i-a (8,11,13,17), i-ā (19), au-u (10), au-ū (12), o-ā (14), ε-a (15), o-a (16)
Second syllable: ø-zero (2,16)

Three vowel correspondences occur more than once. The most prevalent of these is /i/ in English and /a/ in Hindi. There is also a correspondence between English /æ/ and Hindi /ā/. Finally, there is a correspondence between English /æ/ in final position and the lack of a vowel in Hindi.

4.7.3 ETYMOLOGICAL CONSIDERATIONS

As in the other cases, let us examine the details of the etymologies. These have been grouped by borrowing, chance, and genetic relationship.

**Borrowing** (may also reflect actual genetic relationship):

**DUST: DHUL(1)** Dust is from Germanic *d(w)unst-, which is further related to Sanskrit dhvanis. The Hindi word is a direct borrowing from Sanskrit dhūliḥ, which is related to the English word (cf. Lithuanian dūlis ‘fog’).

**NOSE: NĀSĀ** The Indo-European designation for this body part is reflected in both English (Old English nosu/nasu) and Hindi (Sanskrit nāsā).

**THIN: TANU** (things) Both English (Old English pyynne) and Hindi (Sanskrit tanu) words go back to Indo-European *tenu-s.*
Chance:

**Boil** (v.): **Ubālā** The English verb stems from a loan from Latin *bullīre* (a derivative of *bullāre*, from *bulla* 'bubble'). The source of the Hindi word is unclear, although Hock suggests it may stem from Sanskrit *ud-vṛt- 'to rise up'. There is also the possibility that both the Latin and Hindi words are onomatopoeic.

**Chant**: **Gānā** The English verb is a **borrowing** from French *chanter* 'to chant, to sing', with its ultimate source Latin *cantāre* 'to chant, to sing' (in turn derived from *canēre*, 'to sing, to play (an instrument)'). The source of the Hindi word is a Sanskrit root *gā*- 'to speak, to sing', which Mayrhofer suggests may be related to Indo-European (cf. Lithuanian *giedōti* 'to sing', Old Church Slavonic *gajati* 'to crow').

**Chest** (body): **Chātī** The English word stems from the Latin *cista* 'chest, box', which is in turn derived from Greek *kístē*, 'chest, box'. The source of the Hindi word is unclear, but it may stem from Sanskrit *chādati* 'to cover'. In any case, the similarities seem to be due to chance.

**Chew**: **Cābānā** The English word is related to other Indo-European words (cf. German *kauen*, Russian *žvat*') and is from the IE root *g[ī]eu*- 'chew'. The Hindi word stems from Sanskrit *carv-*, but the further etymology is unclear, leading to the possibility that the Sanskrit word does not reflect an Indo-European root.

**Far**: **Dūr** The English word *far* (Old English *fear*) reflects a widespread Germanic word (cf. German *fern* 'far'). This root is fur-
ther related to Sanskrit *paras*. The source for Hindi, however, is Sanskrit *dūrah*, which, while of Indo-European origin, is not related to the English word.

**FEAR:** The Modern English *fear* (Old English *fær*) is the descendant of a root related to ‘travel’ (cf. German *fahren* ‘to travel’), supposedly due to the danger (Latin *perículum*, German *Gefahr*) of travel in former times. The Hindi word was not available in my etymological sources. According to Hock, Hindi words with an initial retroflex are rarely very old.

**GRASS:** The Indo-European word for ‘to grow, to become green’ *ghre*- is at the basis of English *grass*. Grass is thus the freshly green growth on the ground. The source of the Hindi word is most likely Sanskrit *ghāsatī* ‘eats’, unrelated to the English word.

**HOUND:** Germanic *hundaz* is reflected in the English word, which ultimately goes back to Indo-European *k'ũuð*[n]. The Sanskrit word for ‘dog, hound’ was *śvan-*, which is not reflected in Hindi *kutta*. The Hindi word is most likely an onomatopoeic item, which developed after the Sanskrit period.

**LITTLE:** English *little* (Old English *lītel*) comes from the Germanic root *leut-* ‘to stoop’. The Hindi word is a borrowing from Sanskrit. The word in Sanskrit is probably related to English *light* ‘not heavy’.

**NEAR:** English *near* (Old English *nīg)n* reflects Germanic *nāhwod-*. The Hindi form is a borrowing from Persian
(Indo-Iranian *nazard*). Because Hindi has borrowed its word, this correspondence cannot be used as evidence of genetic relationship.

**RED: RAKT** The English word (Old English *read*) reflects Germanic *raudaz*, which is ultimately derived from Indo-European *rudhro-*. The source of the Hindi word is Sanskrit *rakta-īråjaḥ* ‘colored, red’. The continuation of the Indo-European root is reflected in Sanskrit *rudhirāḥ* ‘red, bloody’, so one can see that the resemblance between Hindi and English in this case is chance.

**SHOULDER: SKANDH** The exact etymology of English shoulder is unclear, but it is attested only in West Germanic. Some etymological dictionaries attempt to tie the word to the Indo-European root *̞skel-* ‘to cut’, as animal bones were used as implements (shovels) in ancient times. Likewise somewhat unclear is Sanskrit *skandha-*, the source (via borrowing) of the Hindi word. Perhaps it comes from *skándhaḥ* ‘twig, branch’, but in any case, the Hindi and English words are not related.

**WISH (v.): ĀṢĀH karnā** English *wish* reflects Germanic *wunskjan*, whose the Sanskrit counterpart was *vānch* (cf. *van* ‘to ask’), not reflected in the contemporary Hindi form. Both the Germanic and Sanskrit forms probably go back to the Indo-European root *wen-* ‘search’ (cf. Latin *venus* ‘longing for love’, Sanskrit *vanas-* ‘longing’). The source of the Hindi word, however, is Sanskrit *āśāḥ* (later *āśā*), which is built from two roots ā* ‘hither’ and *śāsti* ‘commands, instructs, punishes’. The similarity is thus due to chance.
**Genetic Relationship:**

**BROTHER:** BHĀṬ The immediate protoform for the English word is Germanic *brōper-* and the ultimate form is Indo-European *bhrāter-*, which is also the ultimate form for the Hindi word. The method has scored a success.

**MOUSE:** MŪSIKĀ (house), MŪS (field) The Indo-European word for the rodent is reflected in both English and Hindi. In Old English, the word was mās, while the word was mūsha in Sanskrit.

**TWO:** DŌ The English form (Old English twā (feminine)) and the Hindi form (Sanskrit dvāu, dvā) stem from an Indo-European form *dwō[u]*.

**YOUNG:** ŠAVĀN English young (Old English (g)iung) reflects Germanic *yun-gāz*, which, along with Sanskrit yuvaśaḥ, yūvan go back to Indo-European *yuwen*.

### 4.8 ENGLISH-HINDI-FINNISH

The three-way correspondence among English, Hindi and Finnish, perhaps the most telling test case for mass comparison, number 10/84, or 11.9%, well above Greenberg’s 2.2% for LIA and significantly above the 8% from 1955.

1. carry  
2. cook (boil)  
3. eye  
4. hut  
5. narrow  

<table>
<thead>
<tr>
<th>English</th>
<th>Hindi</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>carry</td>
<td>kar ‘hand’</td>
<td>kārsiā</td>
</tr>
<tr>
<td>cook (boil)</td>
<td>khaunā</td>
<td>keittää</td>
</tr>
<tr>
<td>eye</td>
<td>ākh</td>
<td>näko</td>
</tr>
</tbody>
</table>
| hut | kōṭi ‘citadel, castle’,  
kōṭhī ‘bungalow, mansion’ | koti |
| narrow | nārā ‘string’, nāṭī ‘vein’ | niukka |
4.8.1 **CONSONANT CORRESPONDENCES AMONG ENGLISH, HINDI AND FINNISH**

Initial: p-ph-p (8), k-k-k (1), k-kh-k (2), h-k-k (4), n-n-n (5,6,7), s-s-s (9), w-v-v (10)

Medial: r-r-r (1), y-kh-k (3), r-r-k (5), v-bh-p (6), l-l-l (9), v-l-k (9)

Final: k-l-t (1), t-t-t (4), l-o-o (6), f-kh-h (8), nd-t-n (10)

The consonant correspondences between English, Hindi, and Finnish are almost picture perfect. Of course, this was the basis for scoring the words as possible cognates in the first place. The problem is that, aside from /n/ in all three languages, no correspondence repeats, leaving us with little evidence from which to draw conclusions.

4.8.2 **VOWEL CORRESPONDENCES AMONG ENGLISH, HINDI AND FINNISH**

First syllable: e-a-ã (1), u-au-ei (2), ai-ã-ã (3), e-o-o (4), e-ã-iu (5), ei-ã-a (6), u-a-uo (7), e-ã-u (8), e-a-y (9), i-ã-ai (10)

Second syllable: i-zero-iã (1), o-ã-a (5), e-i-a (6), ai-i-i (9)

The vowels show no recurrent correspondences across all three languages.
4.8.3 ETYMOLOGICAL CONSIDERATIONS

As before, let us examine the details of the etymologies. These have been grouped by chance, related in two languages (chance in third), borrowing between two languages (third unrelated), borrowing between two languages (third related), and onomatopoeia.

**Chance** (in all three):

**CARRY**: **KAR** 'hand': **KÄRSIÄ** The English word is from Latin *carrus* 'cart', which was probably a borrowing from Celtic (cf. Welsh *car*, Breton *karr*, Irish *carr*, all meaning 'chariot'). The source of the Hindi word is Sanskrit *karāḥ*, the etymology of which is unclear. Perhaps it is related to the root meaning 'to do', thus known as 'the doing part'. The Finnish word has cognates in the other Balto-Fennic languages (cf. Karelian *käršü* 'consolation', Veps *käržün*, Estonian *käršida*), but not elsewhere, and is thus unlikely to be inherited from Proto-Uralic.

**COOK**: **KAULNÄ**: **KEITÄÄ** The English verb *to cook* reflects an early loan from Latin *coquere* 'to cook, to ripen'. The etymology of the Hindi word is unavailable from my sources. The word represented in Finnish is widely attested in Uralic, encompassing both Latin meanings (cf. Estonian *keeta* 'boil', Cheremis *küja* 'ripen, boil (v.t.)', *küljkte* 'ripen, boil (v.i.)', Yurak Samoyed *kuu*-'ripen (of berries)'). Once again, there is no evidence for a genetic relationship.
NARROW: NÄRA: NIUKKA The English word is related to German Narbe ‘scar, drawing together of a wound’ and Lithuanian narys ‘a loop (of thread)’, and perhaps related to the cognates of nerve. Hock suggests that the Hindi word may go back to a Sanskrit word nädāḥ ‘reed’, but Mayrhofer offers no further etymology of the Sanskrit word. Fenno-Ugric cognates for the Finnish word are limited to the closely-related Karelian dialects (cf. Karelian nīukka, Aun. nīukku).

SALIVA: SĀLILA: SYLKI The Hindi word most likely goes back to Sanskrit salilām ‘sea’, which stems from the Indo-European root *sal- ‘salt’. English has borrowed from Latin salīva. It has been suggested that the Latin word is related to Latin salix ‘willow’, but the exact origins remain unclear. The cognates for Finnish sylki are limited to the Fenno-Ugric languages (cf. Lapp čolgā, Cheremis šūwël-wūt (wūt ‘water’)). There appear to be three different roots in operation here, making chance the best explanation for the resemblances.

Related in Hindi and English (chance in Finnish)

EYE: ĖKH: NĀKO The Germanic root *aug- is reflected in English (Old English ēage; cf. also Gothic augō, Danish øje), which ultimately goes back to Indo-European *okus- (cf. Lithuanian akiš, Latin oculus, Old Church Slavonic oko). The Hindi word also goes back to this root (through Sanskrit ākṣi). The Finnish word, along with nāke ‘to see’, is related to words throughout Fenno-Ugric (cf. Hungarian néz ‘see’,
Lapp *niekka* ‘to dream’), but all of the Fenno-Ugric words have an initial *n-* , lacking in the Indo-European root.

NEW: NAYÄ: NUORI English *new* (Old English *nīwe*), as well as Hindi *nāyā* (Sanskrit *navya(s)*) go back to Indo-European *new(i)os*. A root found in Baltic Fennic, but not elsewhere, is reflected in Finnish (cf. Karelian *nuori*, Lude *nuor*, Veps *nor*). Because the Finnish word is found only in Baltic Fennic, it is unlikely that this reflects a word long in existence. Therefore, this is not indicative of genetic relationship.

WIND (n.): VĀT: VAINU The Germanic root *wendoz* is reflected in English *wind*. Both go back to an Indo-European root *wē̄*, ‘to blow’. The IE root also gave rise to Sanskrit *vā* ‘blow’ and *vātas* ‘wind’, the source of the Hindi form (which is actually a borrowing from Sanskrit). The Finnish word was unavailable from my etymological sources, but there is no evidence to suggest that the relationship is more than merely chance.

**Borrowing between two languages** (third unrelated)

HUT: KOTI: KOTI The English word reflects a borrowing from French (*hutte*), which in turn had borrowed from Old High German (*huota*). The word appears to have come ultimately from Indo-European *[s]ku-* ‘to cover’ (cf. *hide* (v.), German *Haut* ‘skin’). The Finnish word was a borrowing from Avestan *kata-* ‘chamber’ (cf. Modern Persian *kad* ‘house’). In addition to Finnish, the word also was taken into Hungarian (*ház*) and Lapp (*koti*). The Hindi word
was likewise borrowed, according to Mayrhofer, from Dravidian. Borrowing in all three instances makes the exact relationship difficult to explain.

**Borrowing between two languages** (third related)

NAVEL: NĀBHI: NAPA  English *navel* goes back to Germanic *nabhlon*, which in turn is from Indo-European. Also from the same Indo-European root is Sanskrit nābhī, whence the Hindi form. Cognates of the Finnish form are found only in the Balto-Fennic group (cf. Karelian, Lude nāba, Livonian nāba, Estonian naba). Collinder suggests a possible Baltic source for the word, thus tying the Finnish word into the Indo-European through borrowing.

**Onomatopoeia**

PUFF: PHŪKHĀ: PUHALTTAA  The English word is of onomatopoeic origin, as, most likely, is the Hindi word. Finnish puhaltaa reflects a widespread Uralic word (cf. Lapp bosso-boso ‘to blow’, Ziryene pušky ‘to blow’, but this root is also probably an onomatopoeic form).

4.9 **DISCUSSION OF CORRESPONDENCES**

What are we to make of such a list of correspondences? As the similarities are well above the 2% margin proposed by Greenberg, are we to conclude that English, Hindi, and Finnish are related to each other? Are we to conclude, based on the evidence presented,
that English and Hindi are only slightly more closely related than Finnish and Hindi?

After applying the mass comparison, followed by the comparative method, I arrived at different percentages for the various groupings. Of the 8.3% correspondence between English and Finnish uncovered by mass comparison, none of the 8.3% was due to actual genetic relationship, 28.6% were attributable to borrowings. Thus, 71.4% of the items scored as (possible) cognates were due to pure chance.

An initial percentage of 19.1% were uncovered for Hindi and Finnish. None of this 19.1% was attributable to actual genetic relationship, 12.5% were due to borrowing, and 6.3% each were due to onomatopoeia and nursery language. In the end, a full 75% of the 19.1% were due to chance.

In the case of actual genetic relationship, between Hindi and English, 23.8% were the result of mass comparison. 20% were due to genetic relationship reflected in English and Hindi. There was an additional 15% that were the result of borrowing between Hindi and Sanskrit, which can, in a manner of speaking, be considered to be genetic relationship between Hindi and English, as Hindi is ultimately derived from Sanskrit. Thus, 35%, far less than half, of the correspondences uncovered by mass comparison were due to genetic relationship. Moreover, it did not uncover an actual relationship, namely English horn and Hindi स्तग, (cf. section 3.1.1).
In the end, 65% of the resemblances selected by the application of mass comparison were due to chance.

Finally, in the resemblances between English, Hindi, and Finnish, originally at 11.9%, resulted in none being reflective of relationship among the three. 40% were pure chance in all three, 30% were indicative of relationship between Hindi and English, 10% are the result of borrowing between two languages, with the third one unrelated, 10% were due to borrowing between two languages, with the third one related, and 10% were the result of onomatopoeia. In other words, the method has scored a complete miss by scoring relationship between the three languages.

As one can see from the evidence, the method does not fare well in this study. In its best performance, only 35% of the correspondences selected by the method were indicative of genetic relationship. The method simply does not provide the reliable results which one needs in order to establish basic genetic relationship, let alone the distant genetic relationship Greenberg (1955 and 1987) is attempting to establish.

5.0 IMPLICATIONS FOR GREENBERG'S APPROACH

This study shows that Greenberg's method of mass comparison does not work. In its best case scenario, the method has been shown to have only about a 50% chance of predicting the correct results. In its worst case, it has predicted relationship between
languages whose exact affiliation is not well known, and those similarities which the method has found are due solely to chance.

In the study presented here, the percentages between Hindi and Finnish (19.1%) and between English and Hindi (23.8%) are remarkably close. However, only one of the two relationships is truly genetic: the relationship between English and Hindi. Of the data presented there, though, only 15.5% (of all the data) were indicative of genetic relationship.

A method that has a 50-50 chance of predicting correct results in its strongest case cannot be relied upon to classify the languages of the world. After applying mass comparison to find possible connections between languages, the comparative method must also be employed to reach even the lowest level on the cline of confidence.

Looking through lists such as those presented here offers linguists an opportunity to make some rough hypotheses as to the relatedness of two or more languages. The method of multilateral comparisons is unable to predict anything more. It can offer nothing conclusive.
## APPENDIX A.

### 84 CONCEPTS SELECTED FROM GREENBERG’S (1987) AMERIND DICTIONARY

<table>
<thead>
<tr>
<th>English</th>
<th>Hindi</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. angry, irate, mad</td>
<td>kruddh, kupit, nārāz</td>
<td>suuttunut, vihainen</td>
</tr>
<tr>
<td>2. arrive, come</td>
<td>pōḥučnā, ānā</td>
<td>saapua</td>
</tr>
<tr>
<td>3. arrow</td>
<td>tār, bān, shar</td>
<td>nuoli</td>
</tr>
<tr>
<td>4. ashes, cinders</td>
<td>rākh, bhasm</td>
<td>tuhka</td>
</tr>
<tr>
<td>5. aunt</td>
<td>phūphī/būā</td>
<td>tāti</td>
</tr>
<tr>
<td></td>
<td>(father’s sister)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kākī/čāčī</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(father’s brother’s wife)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mausī/māsī</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(mother’s sister)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>māmī</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(mother’s brother’s wife)</td>
<td></td>
</tr>
<tr>
<td>6. bathe, wash</td>
<td>nahānā</td>
<td>kylpeä, uida,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kostuttaa, kastaa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kantaa, tuottaa,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kärsiä</td>
</tr>
<tr>
<td>7. bear (v.), carry</td>
<td>ḍhōnā, kar ‘hand’</td>
<td>takana, taakse,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>jäljessä</td>
</tr>
<tr>
<td></td>
<td></td>
<td>puhaltaa, tuulla,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lentää</td>
</tr>
<tr>
<td>8. behind</td>
<td>pīče</td>
<td>kiehua, kuohua,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kiehutta</td>
</tr>
<tr>
<td>9. blow (v.)</td>
<td>phūkhnā</td>
<td>murtaa, rikkoa,</td>
</tr>
<tr>
<td>10. boil (v.)</td>
<td>ubalnā, khaulnā</td>
<td>särkeä, taittaa</td>
</tr>
<tr>
<td>11. break (v.)</td>
<td>tūnā (v.i.), tornā</td>
<td>rinta, etupuoli, sydän</td>
</tr>
<tr>
<td>12. breast</td>
<td>čhātī, sīnā, stan,</td>
<td>tuoda, saada,</td>
</tr>
<tr>
<td></td>
<td>kuč (wmn’s)</td>
<td>aiheuttaa</td>
</tr>
<tr>
<td>13. bring</td>
<td>le ānā, lānā</td>
<td>leveä, laaja, avara,</td>
</tr>
<tr>
<td>14. broad, wide</td>
<td>čaurā</td>
<td>selvä</td>
</tr>
<tr>
<td>15. brother</td>
<td>bhāṛī</td>
<td>veli, lähimminen</td>
</tr>
<tr>
<td>16. burn (v.)</td>
<td>ālānā, ālānā</td>
<td>polttaa, palaa, hehkua</td>
</tr>
<tr>
<td>17. chest (body)</td>
<td>ċhāṭī, sīnā</td>
<td>rinta, rintakehä</td>
</tr>
<tr>
<td>18. chew</td>
<td>ċabānā</td>
<td>pureskella</td>
</tr>
<tr>
<td>19. clean (v.)</td>
<td>sāf karna</td>
<td>puhdistaa, siistiä, kiilloittaa</td>
</tr>
<tr>
<td>20. close (v.)</td>
<td>band karna</td>
<td>sulkea, ummistaa, lopettaa</td>
</tr>
<tr>
<td>21. come</td>
<td>ānā, pōhūčnā</td>
<td>tulla, saapua, johtua</td>
</tr>
<tr>
<td>22. cook (v.)</td>
<td>pakānā</td>
<td>keittiä, valmistaa</td>
</tr>
<tr>
<td>23. cover (v.)</td>
<td>dhaknā, ċaṛhānā</td>
<td>peittää, kattaa, verhota</td>
</tr>
<tr>
<td>24. cry (v.), scream, whine</td>
<td>ċillānā, rōnā, pukārnā</td>
<td>huutaa, huudahtaa, itkeä</td>
</tr>
<tr>
<td>25. dark, black</td>
<td>andherā, kālā (black), dhūḍhīlā, gahrā</td>
<td>pimeä, synkkä, ruskeaihoinen, tietämätön, tumma</td>
</tr>
<tr>
<td>26. dirty, filthy</td>
<td>gandā, mailā</td>
<td>likainen, kurainen</td>
</tr>
<tr>
<td>27. dog, hound</td>
<td>kuttā, kukkur</td>
<td>koiru</td>
</tr>
<tr>
<td>28. dust, powder</td>
<td>dhūl(i), gard, cūṛṇ</td>
<td>tomu, pöly</td>
</tr>
<tr>
<td>29. ear</td>
<td>kān</td>
<td>korva</td>
</tr>
<tr>
<td>30. eye</td>
<td>ākh</td>
<td>silmä, näkö</td>
</tr>
<tr>
<td>31. fall (v.), plunge, drop</td>
<td>girnā, ċharānā</td>
<td>pudota, kaatua</td>
</tr>
<tr>
<td>32. far, distant</td>
<td>dūṛ</td>
<td>kaukana, kauas, etäällä</td>
</tr>
<tr>
<td>33. fear, anxiety</td>
<td>ċar, bhay</td>
<td>pelko</td>
</tr>
<tr>
<td>34. flea</td>
<td>pissū</td>
<td>kirppu</td>
</tr>
<tr>
<td>35. fly (v.)</td>
<td>uṛnā</td>
<td>lentää, kiitää, rientää, lichua</td>
</tr>
<tr>
<td>36. fruit</td>
<td>phal</td>
<td>hedelmä</td>
</tr>
<tr>
<td>37. girl</td>
<td>laṛkī</td>
<td>tyttö, apulainen</td>
</tr>
<tr>
<td>38. grass</td>
<td>ghās</td>
<td>ruoho, nurmi</td>
</tr>
<tr>
<td>39. hair</td>
<td>bāl (head), rōm, lōm, rōyā (body)</td>
<td>hius, tukka, karva, nukka</td>
</tr>
</tbody>
</table>
40. hit (v.)  mārnā  lyödä, iskeä, törmätä, sattua
41. horn  sīg, višān
42. house  ghar, makān, sadan, bhavan, kōṭ 'citadel, castle', kōṭhā 'large room, warehouse', kōṭīṇ 'bungalow, mansion'
43. husband  pati
44. kill  mār ḍālnā  aviomies  tappaa, surmata, tehdä
45. know  jānnā, mālūm hōnā
46. laugh (v.)  hāsṇā
47. leg  tāg, pair
48. light (n.)  prakāśh, ḫārī, rōshnī, dīpak (lamp), battī (lamp)  tietää, tuntea, osata  nauroma  sääri, jalka  valo
49. lip  (h)ōṭh, ṣṭh
50. liver  kalēṛa, ḫagār
51. mouse  cuhiyā, mūsikā (house), mūs (field)
a. bat  gādur, ḫamgādar  huuli, reuna, parras  maksa  hiiri
52. narrow  saṅkīṛn, tag,
53. navel, bellybutton  nāṭā 'string', nāṭī 'vein', nābhi, ḍhōḍhē, tundī
apa
54. near  nikaṭ, pās, nazdīk(ī)
55. neck  gardan, grīvā
56. nose  nāk, nāśā
57. old, ancient  būṛhā (person), purānā, prācīn
tuska, kipu, kärsimys
58. pain  pīṭhā, dard
59. penis  shishn, līṅ
60. person  vyakti (indiv'1), mānava (human)
henkilö, persoona
61. rain  bārīsh, varṣā, vṛiṣṭi  sade
<table>
<thead>
<tr>
<th>English</th>
<th>Tamil</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>62. red</td>
<td>lāl, sūrā, aruṇ, rakt, lōhit, shōnīt</td>
<td>punainen</td>
</tr>
<tr>
<td>63. root</td>
<td>jāṛ, mūḷ, kand (edible only)</td>
<td>juuri, alku</td>
</tr>
<tr>
<td>64. round</td>
<td>gōl</td>
<td>pyöreä, runsas</td>
</tr>
<tr>
<td>65. saliva, spit</td>
<td>lār, rāl, lālā</td>
<td>sylki</td>
</tr>
<tr>
<td>66. seek, look</td>
<td>salil 'water', sāl 'damp'</td>
<td>hakea, etsiä, tavoitella, pyytää, koettaa</td>
</tr>
<tr>
<td>search</td>
<td>khōjīnā, dhūrīnā</td>
<td>loistaa, paistaa, lyhyt, vajaa</td>
</tr>
<tr>
<td>67. shine</td>
<td>patā lagānā, chān dālnā, dēkh dālnā, talāshna</td>
<td>hartia, olkapää, laulaa</td>
</tr>
<tr>
<td>68. short</td>
<td>čamaknā, čamkānā</td>
<td>pieni, vähäinen, kapea, heikko, mitätön, hapan, äreä, nyrpeä, vahva, voimakas, luja, mieslukuinen, väkeva niellä, nielaista kives</td>
</tr>
<tr>
<td>69. shoulder</td>
<td>kiiltāā, hohtaā</td>
<td></td>
</tr>
<tr>
<td>70. sing, chant</td>
<td>chōtā, laghu, nātā, niukka, riittämätön thīgnā</td>
<td></td>
</tr>
<tr>
<td>71. small, little</td>
<td>kandhā, skandhā, as ulkonema</td>
<td></td>
</tr>
<tr>
<td>72. sour</td>
<td>gānā</td>
<td>laulaa</td>
</tr>
<tr>
<td>73. strong, muscular tough</td>
<td>čhōtā, laghu, nāhā, thōrā, alpa, kam (amount)</td>
<td>pieni, vähäinen, kapea, heikko, mitätön, hapan, äreä, nyrpeä, vahva, voimakas, luja, mieslukuinen, väkeva niellä, nielaista kives</td>
</tr>
<tr>
<td>74. swallow (v.)</td>
<td>khattā, amlā</td>
<td></td>
</tr>
<tr>
<td>75. testicle, ball</td>
<td>ḫṛṣṭ-puṣṭ, ḫattā-kaṭṭā, tagṇā, balī (persons)</td>
<td>ohut, hieno, laiha, mieto</td>
</tr>
<tr>
<td>76. thin</td>
<td>mazbūt, pakkā, ṭīkāū</td>
<td>kaulanetuosa, kurkku</td>
</tr>
<tr>
<td>77. throat</td>
<td>nīgalnā, līknā</td>
<td></td>
</tr>
<tr>
<td>78. two</td>
<td>anḍ-granthī, anḍ (egg) ṭrṣaṇ</td>
<td></td>
</tr>
<tr>
<td>79. vagina</td>
<td>patlā, bārīk, mahīn, tanu (things), dublā, dublā-patlā (people)</td>
<td></td>
</tr>
<tr>
<td>80. wind (n.)</td>
<td>galā, kaṇṭh, grāsīkā, halaq</td>
<td></td>
</tr>
<tr>
<td>81. wing</td>
<td>ohut, hieno, laiha, mieto</td>
<td></td>
</tr>
<tr>
<td>82. wish (v.)</td>
<td>do</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pākh, pakṣ, par, ḍainā</td>
<td>kaksi, molemmat</td>
</tr>
<tr>
<td></td>
<td>cahanā, iččā karnā,</td>
<td>emäin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tuuli, hengitys, vainu, vihi</td>
</tr>
<tr>
<td></td>
<td></td>
<td>siipi</td>
</tr>
<tr>
<td></td>
<td></td>
<td>toivoa, haluta,</td>
</tr>
</tbody>
</table>
80

hope
83. yesterday
84. young, new

āśhā karnā
kal, hāl mē
čhōtā, kamsin, taruṇ,
tāzā

toivottaa
eilen
nuori, vasta-alkava,
kishōr, ḫvān, yuvā,
kokematon
Sources


Bloomfield, Leonard. 1925. On the sound system of Central Algonquian. Lg. 1.130-56.


Campbell, Lyle. 1988. Review article on Greenberg 1987. Lg. 64.591-615.


Matisoff, James A. 1990. On megalocomparison. Lg. 66.106-120.


